THE AMERICAN MATHEMATICAL
MONTHLY

Volume 106, Number 9 November 1999

NOTICE TO AUTHORS

The MONTHLY publishes articles, as well as notes and other features, about mathematics and the profession. Its readers span a broad spectrum of mathematical interests, and include professional mathematicians as well as students of mathematics at all collegiate levels. Authors are invited to submit articles and notes that bring interesting mathematical ideas to a wide audience of MONTHLY readers.

The MONTHLY's readers expect a high standard of exposition; they expect articles to inform, stimulate, challenge, enlighten, and even entertain. MONTHLY articles are meant to be read, enjoyed, and discussed, rather than just archived. Articles may be expositions of old or new results, historical or biographical essays, speculations or definitive treatments, broad developments, or explorations of a single application. Novelty and generality are far less important than clarity of exposition and broad appeal. Appropriate figures, diagrams, and photographs are encouraged.

Notes are short, sharply focussed, and possibly informal. They are often gems that provide a new proof of an old theorem, a novel presentation of a familiar theme, or a lively discussion of a single issue.

Articles and Notes should be sent to the Editor:

ROGER A. HORN
1515 Mineral Square, Room 142
University of Utah
Salt Lake City, UT 84112

Please send your email address and 3 copies of the complete manuscript (including all figures with captions and lettering), typewritten on only one side of the paper. In addition, send one original copy of all figures without lettering, drawn carefully in black ink on separate sheets of paper. Authors who use LaTeX are urged to use article.sty and its standard environments with no custom formatting

Letters to the Editor on any topic are invited; please send to the MONTHLY's Utah office. Comments, criticisms, and suggestions for making the MONTHLY more lively, entertaining, and informative are welcome.

See the MONTHLY section of MAA Online for current information such as contents of issues and descriptive summaries of forthcoming articles:

http:// www.maa.org /

Proposed problems or solutions should be sent to:

DANIEL ULLMAN, MONTHLY Problems
Department of Mathematics
The George Washington University
2201 G Street, NW, Room 428A
Washington, DC 20052

Please send 2 copies of all problems/solutions material, typewritten on only one side of the paper.

Reprint permission:
DONALD ALBERS, Director of Publications

Advertising Correspondence:
Dave Riska, *driska@maa.org*

Change of address, missing issues inquiries, and other subscription correspondence:
MAA Service Center, *maahq@maa.org*

All at the address:

The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, DC 20036

Recent copies of the MONTHLY are available for purchase through the MAA Service Center,
maahq@maa.org, 1-800-331-1622

Microfilm Editions: University Microfilms International, Serial Bid coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

A Modern Treatment of the 15 Puzzle

Aaron F. Archer

1. INTRODUCTION. In the 1870's the impish puzzlemaker Sam Loyd caused quite a stir in the United States, Britain, and Europe with his now-famous 15-puzzle. In its original form, the puzzle consists of fifteen square blocks numbered 1 through 15 but otherwise identical and a square tray large enough to accommodate 16 blocks. The 15 blocks are placed in the tray as shown in Figure 1, with the lower right corner left empty. A legal move consists of sliding a block adjacent to the empty space into the empty space. Thus, from the starting placement, block 12 or 15 may be slid into the empty space. The object of the puzzle is to use a sequence of legal moves to switch the positions of blocks 14 and 15 while returning all other blocks to their original positions.

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	

Figure 1. The starting position for the 15-puzzle. The shaded square is left empty.

Loyd writes of how he "drove the entire world crazy," and that "A prize of $1,000, offered for the first correct solution to the problem, has never been claimed, although there are thousands of persons who say they performed the required feat." He continues,

> People became infatuated with the puzzle and ludicrous tales are told of shopkeepers who neglected to open their stores; of a distinguished clergyman who stood under a street lamp all through a wintry night trying to recall the way he had performed the feat.... Pilots are said to have wrecked their ships, and engineers rush their trains past stations. A famous Baltimore editor tells how he went for his noon lunch and was discovered by his frantic staff long past midnight pushing little pieces of pie around on a plate! **[9]**

The reason for this hysteria, of course, is that Loyd's puzzle has no solution. Each move causes a transposition of the 16 blocks (where the empty square is considered to contain a blank block), and for the blank to end up in the lower right

corner requires an even number of moves, so the resulting permutation is even. But the desired end placement is an odd permutation of the original, and is hence unobtainable. One must assume Sam Loyd knew this, and from there one can only conjecture how much amusement he derived from driving the American public insane.

The puzzle has inspired a sizable number of articles and references in the mathematical literature. The first of these is a pair of articles published in the *American Journal of Mathematics* in 1879 by W. W. Johnson [**7**] and W. E. Story [**13**]. Johnson's article is an explanation of why odd permutations of the puzzle are impossible to obtain, while Story's article proves that all even permutations are possible. The editors were apparently so apprehensive and defensive about publishing articles on what some might charge to be a frivolous topic that they attached the following justification to the end of Story's article:

> The "15" puzzle for the last few weeks has been prominently before the American public, and may safely be said to have engaged the attention of nine out of ten persons of both sexes and of all ages and conditions of the community. But this would not have weighed with the editors to induce them to insert articles upon such a subject in the American Journal of Mathematics, but for the fact that the principle of the game has its root in what all mathematicians of the present day are aware constitutes the most subtle and characteristic conception of modern algebra, viz: the law of dichotomy applicable to the separation of the terms of every complete system of permutations into two natural and indefeasible groups, a law of the inner world of thought, which may be said to prefigure the polar relation of left and right-handed screws, or of objects in space and their reflexions in a mirror. Accordingly the editors have thought that they would be doing no disservice to their science, but rather promoting its interests by exhibiting this *à priori* polar law under a concrete form, through the medium of a game which has taken so strong a hold upon the thought of the country that it may almost be said to have risen to the importance of a national institution. Whoever has made himself master of it may fairly be said to have taken his first lesson in the theory of determinants. [**13**, p. 404]

The puzzle is a popular topic for books on recreational mathematics or mathematical potpourri, such as [**1**], [**2**], [**4**], [**5**], [**9**], and [**12**], most of which use it as an example to illustrate the consequences of even and odd permutations, as does [**14**]. Various sources have suggested variants of the 15-puzzle, including [**3**], [**4**], [**6**], [**8**], [**10**], and [**15**]. Today the puzzle appears on some computer screen savers, and a version is distributed with every Macintosh computer.

Most references to the 15-puzzle explain the impossibility of obtaining odd permutations and many state Story's result that every even permutation is indeed possible, but this author found only three proofs. R. M. Wilson [**15**] published a more general result in 1974, which we discuss at the end of this article. Ball and Coxeter's book [**1**] refers to [**10**] for a proof, but the article does not fulfill the promise. The arcane terminology of Story's article [**13**] renders it difficult to wade through, and of course it does not take advantage of modern notation developed since then. Spitznagel [**11**] published a proof in 1967, but later wrote that "Over the years there have been published a number of unnecessarily complicated explanations of the puzzle. I confess that I myself once published one of these overly complicated accounts" [**12**]. Indeed, Herstein and Kaplansky [**5**] write that

"No really easy proof seems to be known." This article intends to rectify that deficiency.

2. SOLUTION. It should be noted that the proof provided here was developed independently of the previous proofs, but coincidentally shares some ideas with Story's proof [**13**].

We call each of the 15 pieces *blocks*, and the 16 different squares on the board we call *cells*. For reasons that soon become apparent, we number the cells in the snakelike pattern shown in Figure 2. We can think of the empty cell as being occupied by a *blank block*. Each legal move then consists of "moving the blank," that is, exchanging the blank block with one of its horizontal or vertical neighbors. A *placement* is a bijection from the set of blocks (including the blank) to the set of cells—in other words, a snapshot of the board between moves. Given an initial placement, we wish to determine what other placements are attainable through a sequence of legal moves.

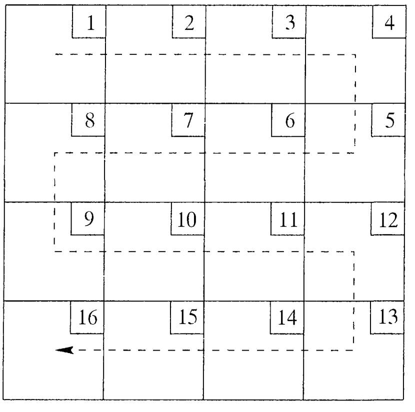


Figure 2. The dashed line and the numbers in the corner of each cell indicate a special ordering of the cells that we use to define equivalence classes of placements.

Notice that by moving the blank block along the snaking path of Figure 2 we can move the blank to any cell without changing the order of the remaining blocks along this path. This leads us to define an equivalence relation on the set of placements, two placements being equivalent if we can obtain one from the other by moving the blank along the snaking path. Each equivalence class is called a *configuration*, and contains 16 placements, one for each cell the blank can occupy. If block i occupies cell j and the blank occupies a higher numbered cell, then we say block i is in *slot* j; otherwise it is in slot $(j-1)$. Refer to Figure 3 for an example. All placements in a given configuration have the 15 blocks in the same slots, so we can denote a configuration by $[a_1, \ldots, a_{15}]$, where a_i is the slot that block i occupies in the configuration.

Every move of the blank block effects a permutation on the slots occupied by the blocks. For example, moving the blank from cell 10 to cell 15 causes the permutation $(10, 11, 12, 13, 14)$ because the block originally in cell 15 (slot 14) is moved to cell 10 (which becomes slot 10) and the blocks in cells 11 through 14 are bumped up one slot. A configuration $[a_1, \ldots, a_{15}]$ subjected to the permutation σ is transformed into the configuration $[a_1, \ldots, a_{15}]\sigma = [a_1\sigma, \ldots, a_{15}\sigma]$; since our permutations act on the right, we multiply them left to right. See Figure 3 for an example.

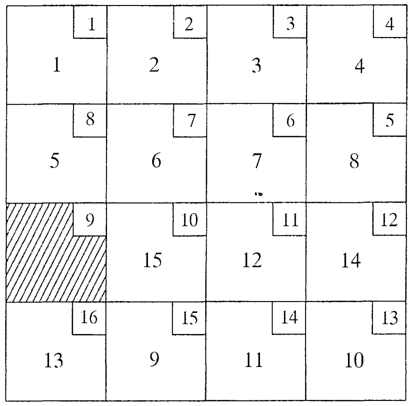


Figure 3. The placement shown here corresponds to the configuration $C = [1, 2, 3, 4, 8, 7, 6, 5, 14, 12, 13, 10, 15, 11, 9]$. Since the initial placement of Figure 1 corresponds to the configuration $I = [1, 2, 3, 4, 8, 7, 6, 5, 9, 10, 11, 12, 15, 14, 13]$, subjecting the initial configuration to the permutation $\sigma = (9, 14, 11, 13)(10, 12)$ yields C. This is an even permutation, so by Theorem 3, C is obtainable from I.

Let $\sigma_{i,j}$ denote the permutation achieved by moving the blank from cell i to cell j. Then clearly $\sigma_{i,i+1}$ is the identity, and $\sigma_{j,i} = \sigma_{i,j}^{-1}$. This leaves 9 permutations for us to work out. These are tabulated in Table 1. The key point is that one can move the blank along the snaking path of Figure 2 to any cell without changing the configuration. Therefore, the first nine permutations listed in Table 1 and their inverses may be applied *in any order*, so the problem reduces to identifying the subgroup of S_{15} (the symmetric group on the 15 slots) generated by these permutations. We prove that these permutations generate A_{15} (all even permutations).

TABLE 1. A summary of all possible permutations of slots attained by moving the blank block. Moving the blank from cell i to cell j effects the permutation $\sigma_{i,j}$.

$\sigma_{1,8}$	$= (1, 2, 3, 4, 5, 6, 7)$
$\sigma_{2,7}$	$= (2, 3, 4, 5, 6)$
$\sigma_{3,6}$	$= (3, 4, 5)$
$\sigma_{5,12}$	$= (5, 6, 7, 8, 9, 10, 11)$
$\sigma_{6,11}$	$= (6, 7, 8, 9, 10)$
$\sigma_{7,10}$	$= (7, 8, 9)$
$\sigma_{9,16}$	$= (9, 10, 11, 12, 13, 14, 15)$
$\sigma_{10,15}$	$= (10, 11, 12, 13, 14)$
$\sigma_{11,14}$	$= (11, 12, 13)$
$\sigma_{n,n+1}$	$= id, n = 1, 2, \ldots, 15$
$\sigma_{i,j}$	$= \sigma_{j,i}^{-1}$ for all relevant $i > j$

Lemma 1. *For $n \geq 3$ the 3-cycles generate A_n.*

Proof: By definition, all elements of A_n can be written as a product of an even number of transpositions. If a, b, c, and d are distinct, then $(a, b)(c, d) = (a, b, c)(a, d, c)$, $(a, b)(b, c) = (a, c, b)$, and $(a, b)(a, b) = id$. ∎

For $n \geq 5$, Lemma 1 also follows directly from the fact that A_n is simple, since the set of 3-cycles is closed under conjugation. Let us call a 3-cycle *consecutive* if it is of the form $(k, k + 1, k + 2)$.

Lemma 2. *For* $n \geq 3$, *the consecutive* 3-*cycles* $\{(1, 2, 3), (2, 3, 4), \dots, (n-2, n-1, n)\}$ *generate* A_n.

Proof: Since the 3-cycles generate A_n, it suffices to show that the consecutive 3-cycles generate all 3-cycles. This is trivial for $n = 3$. For $n \geq 4$ we see by induction that we can generate all 3-cycles not containing both 1 and n. To generate $(1, x, n)$, let $y \in \{1, \dots, n\} \setminus \{1, x, n\}$. Then $(1, x, n) = (y, x, n)(1, x, y)$. Of course, $(1, n, x) = (1, x, n)^2$. ∎

Theorem 3. *The cycles listed in Table* 1 *generate* A_{15}.

Proof: Since all the cycles are odd, they are even permutations, so they generate a subgroup of A_{15}. Note that for any permutation σ we have $\sigma^{-1}(a_1, \dots, a_k)\sigma = (a_1\sigma, \dots, a_k\sigma)$. Thus,

$$(1,2,\dots,7)^{-n}(3,4,5)(1,2,\dots,7)^{n} \text{ yields } (1,2,3),\dots,(5,6,7);$$

$$(5,6,\dots,11)^{-n}(7,8,9)(5,6,\dots,11)^{n} \text{ yields } (5,6,7),\dots,(9,10,11) \text{ and}$$

$$(9,10,\dots,15)^{-n}(11,12,13)(9,10,\dots,15)^{n} \text{ yields } (9,10,11),\dots,(13,14,15)$$

as n assumes the values $-2, -1, 0, 1$, and 2. This constitutes all consecutive 3-cycles in S_{15}, so by Lemma 2 it generates A_{15}. ∎

Thus, given any two placements Pl_1 and Pl_2 belonging to configurations Cf_1 and Cf_2, respectively, Pl_2 is obtainable from Pl_1 if and only if Cf_2 is an even permutation of Cf_1. Stated directly in terms of the placements, we see that if Pl_1 and Pl_2 have the blank in the same cell then Pl_2 is obtainable from Pl_1 if and only if Pl_2 is an even permutation of the 15 numbered blocks in Pl_1. Let n be the number of moves the blank cell in Pl_1 is away from the blank cell in Pl_2. Since each move of the blank block causes a transposition of two blocks, then for n odd (respectively even) Pl_2 is obtainable from Pl_1 if and only if Pl_2 is an odd (respectively even) permutation of the 16 blocks in Pl_1.

3. GENERALIZATIONS. What follows is, in some sense, the broadest generalization of the 15-puzzle. Given any connected graph on n vertices, we can label the vertices with n labels, one of which we call the *blank label*. Each move consists of interchanging the blank label with the label on an adjacent vertex. We then ask which of the $n!$ labelings may be obtained from a given initial labeling through a sequence of moves. More precisely, we ask what permutations of the $(n-1)$ ordinary labels (a subgroup of S_{n-1}) can be obtained by a sequence of moves that returns the blank to its original vertex v (since the subgroups obtained for different choices of v are isomorphic). The 15-puzzle is a special instance of this, corresponding to the graph $P_4 \times P_4$ (the cartesian product of the path on four vertices with itself) depicted in Figure 4. The vertices correspond to cells, the labels (not depicted) correspond to blocks, and the edges show which cells are adjacent.

The crux of the method presented in Section 2 lies in inducing equivalence classes and defining slots by the position of the blank along a hamiltonian path (a path that visits every vertex of the graph exactly once). The method is applicable to any graph containing a hamiltonian path, using any such path. Thus, for the 15-puzzle we could have used a spiral instead of the serpentine pattern of Figure 2. Another example is the Petersen graph. Numbering the vertices as in Figure 5, we

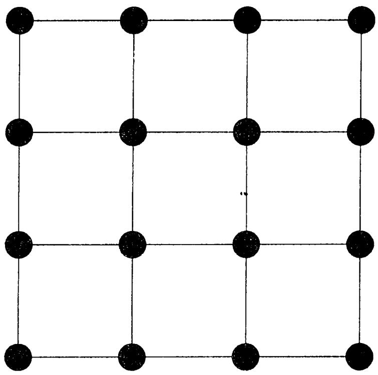


Figure 4. The graph $P_4 \times P_4$.

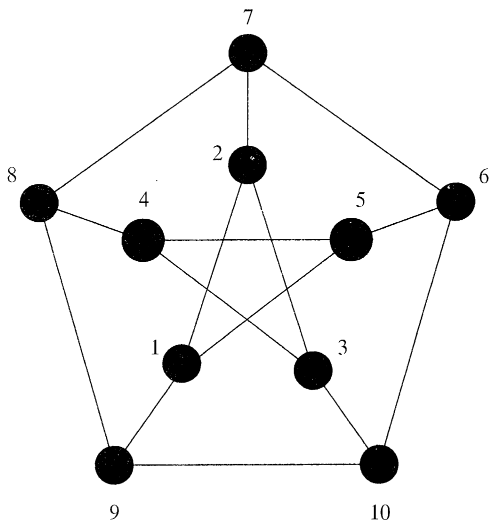


Figure 5. For the famous Petersen graph, each labeling is obtainable from every other by a sequence of legal moves. The vertices are numbered to indicate a hamiltonian path.

see that our desired group is generated by $\sigma_{1,9}$, $\sigma_{1,5}$, $\sigma_{2,7}$, $\sigma_{3,10}$, $\sigma_{4,8}$, and $\sigma_{6,10}$, where $\sigma_{i,j} = (i, i+1, \ldots, j-1)$ is the permutation of slots effected by moving the blank label from vertex i to vertex j. Some calculation shows that the group generated is all of S_9; **[15]** explains why this is no coincidence.

We now discuss the general case, where the graph may or may not contain a hamiltonian path. If the graph contains a cut vertex v then none of the labels other than the blank may be moved across v, so the problem decomposes into two parts. Thus, it suffices to consider graphs containing no cut vertices.

In **[15]**, R. M. Wilson solves this case completely. Wilson's amazing result is that with the exception of cycles C_n and the graph θ_0 depicted in Figure 6, the group contains A_{n-1}. Clearly the group contains an odd permutation if and only if the graph contains an odd cycle, that is, the graph is not bipartite. So for bipartite graphs the group is exactly A_{n-1}, and otherwise it is all of S_{n-1}. Thus, aside from the two exceptional cases, either exactly *half* or *all* of the $n!$ labelings are obtainable, depending on whether or not the graph is bipartite. For θ_0, the desired

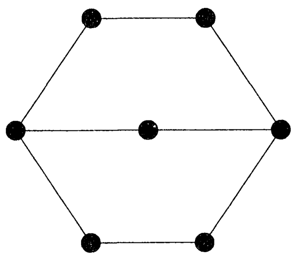


Figure 6. The graph θ_0.

group is $PGL_2(\mathbb{Z}/5\mathbb{Z})$ acting on the projective line over $\mathbb{Z}/5\mathbb{Z}$ (a group of order 120 acting 3-transitively on a set of six elements), yielding six inequivalent labelings. For C_n, the group is $\langle(1, 2, \ldots, n-1)\rangle$, yielding $(n-2)!$ inequivalent labelings. The existence of such a simple complete characterization is surprising. However, Wilson's proof, while elegant, requires considerably more sophisticated mathematics than the simple and elementary proof provided here for the special case of the 15-puzzle.

ACKNOWLEDGMENTS. The author thanks Professors Alan J. Goldman and Arthur T. Benjamin for bringing this problem to his attention, and the latter for many helpful suggestions. Author supported by the Fannie and John Hertz Foundation.

REFERENCES

1. W. W. R. Ball and H. S. M. Coxeter, *Mathematical Recreations and Essays*, 12th ed., U. of Toronto Press, Toronto and Buffalo, 1974, pp. 313–316.
2. J. D. Beasley, *The Mathematics of Games*, Oxford U. Press, Oxford and New York, 1989, pp. 80–81.
3. A. L. Davies, Rotating the fifteen puzzle, *Math. Gazette* **54** (1970) 237–240.
4. M. Gardner, *Martin Gardner's Sixth Book of Mathematical Diversions from Scientific American*, U. of Chicago Press, Chicago, 1971, pp. 64–70.
5. I. N. Herstein and I. Kaplansky, *Matters Mathematical*, Chelsea, New York, 1978, pp. 114–115.
6. S. Hurd and D. Trautman, The knight's tour on the 15-puzzle, *Math. Mag.* **66** (1993) 159–166.
7. W. W. Johnson, Note on the "15" puzzle, *Amer. J. Math.* **2** (1879) 397–399.
8. H. Liebeck, Some generalizations of the 14–15 puzzle, *Math. Mag.* **44** (1971) 185–189.
9. S. Loyd, *Mathematical Puzzles of Sam Loyd: Selected and Edited by Martin Gardner*, Dover, New York, 1959, pp. 19–20.
10. H. V. Mallison, An array of squares, *Math. Gazette* **24** (1940) 119–121.
11. E. L. Spitznagel, Jr., A new look at the fifteen puzzle, *Math. Mag.* **40** (1967) 171–174.
12. E. L. Spitznagel, Jr., *Selected Topics in Mathematics*, Holt, Rinehart and Winston, New York, 1971, pp. 143–148.
13. W. E. Story, Note on the "15" puzzle, *Amer. J. Math.* **2** (1879) 399–404.
14. F. J. W. Whipple, The sign of a term in the expansion of a determinant, *Math. Gazette* **13** (1926) 126.
15. R. M. Wilson, Graph puzzles, homotopy, and the alternating group, *J. Combin. Theory* (Series B) **16** (1974) 86–96.

AARON ARCHER earned his B.S. in mathematics from Harvey Mudd College in 1998, where his research in chromatic graph theory earned him Honorable Mention for the Morgan Prize (AMS/MAA/SIAM). An alumnus of both the Hampshire College Summer Studies in Mathematics and the Budapest Semesters in Mathematics, his time in Hungary inspired him to write an online restaurant guide to Budapest. Aaron is now a Hertz Fellow working toward his Ph.D in operations research at Cornell University. His current research interests include combinatorial optimization and approximation algorithms.
Operations Research Department, Cornell University, Ithaca, NY 14853
Harvey Mudd College, Claremont, CA 91711
aarcher@orie.cornell.edu

Effects of Calculus Reform: Local and National

James F. Hurley, Uwe Koehn, and Susan L. Ganter

1. INTRODUCTION. Over the past decade, the large-scale effort to reform calculus instruction has been a prominent feature of mathematics education. It arose as an energetic response to criticism of the calculus curriculum that culminated in the 1987 conference *Calculus for a New Century*. Substantial support from the National Science Foundation stimulated development and implementation of new calculus curricula at many institutions, including the University of Connecticut.

The instructional practices of calculus-reform programs differ markedly from those that had persisted for decades (some would say, centuries). It is only natural for faculty to question whether the new modes *really* improve the approach that in their own education worked successfully. Some observe little if any improvement in conceptual understanding among students from reform courses, and even complain of *lessened* ability to use computational techniques. They contend (for instance, see [**29**]) that much of what passes for reform really amounts to "dumbing down" a formerly demanding but honest grounding in the power of calculus to a sloppy, imprecise, even misleading shadow of the true nature of the subject. Reformers counter that rote "plug-and-chug" hand symbolic calculation is as intellectually stultifying to teach as to learn, and that their teamwork-fostering technological tools and compelling connections to a broad spectrum of "real-life" issues motivate students to become active learners. They assert that their students emerge from calculus with superior understanding and greater capacity to use its methods successfully, and hence better prepared to complete degrees in mathematics, science, and engineering.

Passionate advocacy of such positions can make for lively lunch-room entertainment, but provides little objective basis for shaping an optimal calculus curriculum. The present paper examines the impact of calculus reform at one medium-sized state university. It also considers the effect of reform at several other institutions. It concludes with a discussion of broader national implications of those outcomes.

The following section describes the features of Connecticut's reform project and includes some representative examples of its activity. For more details, consult [**14**], [**15**], and the instructor guides for the latter [**16**]. Ensuing sections discuss data from common final examinations at Connecticut, results of a five-year longitudinal study of students who took traditional and reform versions of calculus during the project's first year, and results from similar analyses at other sites. The concluding section discusses the national picture, and the kind of further studies appropriate to appraising the impact of calculus reform.

2. THE UNIVERSITY OF CONNECTICUT PROJECT. In 1989, with support from the National Science Foundation Instrumentation and Laboratory Improvement Program, the University's Research Foundation, and the State of Connecticut

High-Technology Program, the Department of Mathematics opened a new computer laboratory. This facility made it possible to expand computer integration from the honors sequence to main-track calculus. Such expansion was a major recommendation of the MAA Committee on Calculus Reform and The First Two Years (CRAFTY) after its site visit the previous year. CRAFTY examined the five-year-old approach in the honors course as one early model for reforming calculus instruction. A brief description of the resulting project follows.

One class hour per week transformed to a computer-laboratory period, and an existing problem-discussion period evolved into a group problem-solving session. In both those settings, students work in self-selecting groups of 3-to-5, without referring to books or notes. In each, the instructor provides some guidance to student exploration and problem-analysis. In the problem hour, students attack a range of conceptual and computational questions, many of which are typical examination problems but a significant number of which are more challenging. In the computer laboratory, students pursue investigation of (generally new) topics by using local programs in the way their laboratory manuals suggest. The first semester, those programs are graphical and numerical True BASIC routines. Second-semester activity includes Maple worksheets and Mathematica notebooks. The remaining two weekly meetings resemble traditional lectures, but students have a more active role in discussion of examples. They often work together (with the aid of graphing calculators) to analyze and investigate problems or properties, and suggest how the instructor should proceed.

At the very first laboratory session, students encounter the function f with formula

$$f(x) = \begin{cases} x\sin\dfrac{1}{x} & \text{for } x \neq 0 \\ 0 & \text{for } x = 0 \end{cases}. \tag{1}$$

A zooming program lets them examine the graph of f near the origin, and leads them to conjecture that it has a limit there; later, they use the sandwich theorem for limits to justify that conclusion. In a subsequent problem-solving hour, they work with similar functions for which graphing calculators provide enough information to suggest continuity or discontinuity. Again, they need to apply appropriate theory to support their conclusions.

Local linearity plays a central role in discussion of differentiation, and the graph of the function f with formula (1) indicates its non-differentiability at $x = 0$. A group project asks them to investigate the function g in which $x^2\cos(1/x)$ replaces $x\sin(1/x)$ in (1). Another outgrowth of local linearity is the approach to implicit differentiation, which uses Euler's method to construct both tables of values and approximate graphs of functional equations $F(x, y) = 0$ near a known (initial-value) point (x_0, y_0). This provides a first experience with numerical solution of differential equations, as well as a ready tool for sketching the graph of equations such as $x^{2/3} + y^{2/3} = 1$.

Initial exposure to area and definite integration occurs in the laboratory, through interaction with a graphical/numerical program that presents the interactive screen in Figure 1. The goal is both to foster conceptualization of $\int_a^b f(x)\,dx$ as a limit of sums and to provide a means of computing the area under the graph of a continuous nonnegative function f over an interval $[a, b]$. Numerical integration thus emerges as a natural and effective means for evaluating definite integrals. This also sets the scene for the fundamental theorem as a remarkable link between

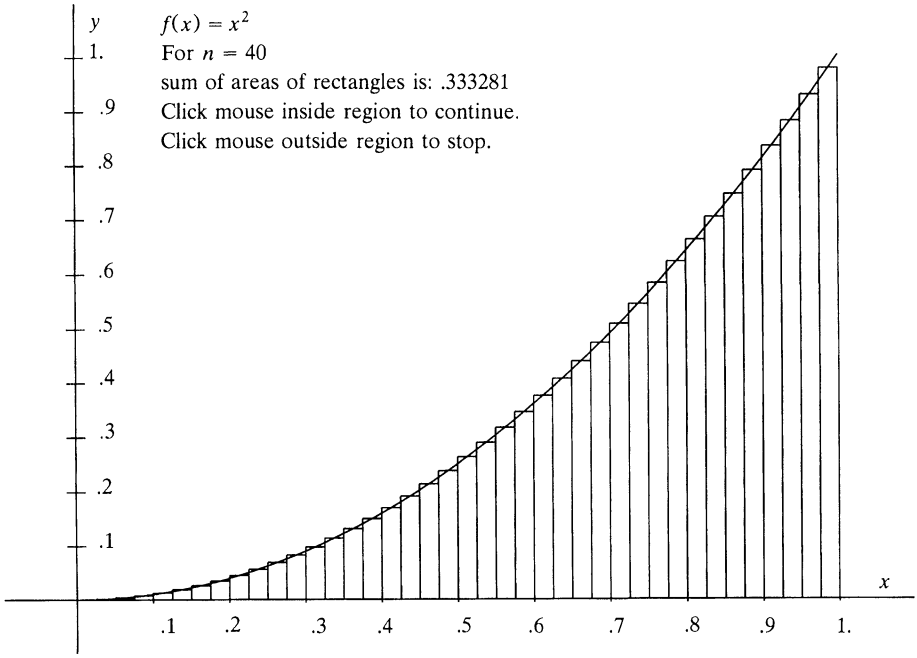


Figure 1

a limiting process and antidifferentiation. Students use numerical summation to estimate $\int_0^1 dx/(1 + \cos x)$, and the corresponding laboratory project considers the accuracy of the right-endpoint approximation $R_n(f)$ and midpoint approximation $M_n(f)$. Students investigate $|I - R_n(f)|/h$, $|I - M_n(f)|/h$, and $|I - M_n(f)|/h^2$, where $I = \int_a^b f(x)\,dx$ is known, and discuss the order-of-magnitude accuracy those ratios suggest.

In the second semester, log-log and semilog plotting illustrate the usefulness of logarithmic and exponential functions in modeling, as most students are actually using those tools in their chemistry laboratory work. Mathematica's `Apart` command removes the algebraic tedium from complex partial-fraction decomposition. Numerical summation appears again, to highlight a perceptible difference between the series $\sum_{n=1}^{\infty} 1/n$ and $\sum_{n=1}^{\infty} 1/n^2$. Laboratory activity explores $s_{2n} - s_n$, and the ensuing project asks for an explanation of *why* for the harmonic series that difference not only does not tend to 0 but in fact seems to converge to $\ln 2$.

Years of teaching experience convinced the project faculty that student perceptions of the importance of material derives in large measure from testing. That prompted them to include conceptual (that is, non-computational) questions regularly not only on discussion worksheets, homework, and group projects but also on quizzes and examinations. For example, the first hour exam in 1989 contained the following true-false questions.

- If $\lim_{x \to c} f(x) = L$ then $\lim_{x \to c^-} f(x) = L$
- If f is continuous at $x = c$ then it must be differentiable there
- If f is continuous on $I = (2, 3)$ then it must have a maximum value on I
- If f is not continuous at $x = c$ then it cannot be differentiable there
- If f is not defined at $x = c$ then $\lim_{x \to c} f(x)$ cannot exist

The common final examination that term showed the graph of a nonnegative function f on a grid over an interval $[a, b]$ and provided five intervals from which the students had to choose the one containing $\int_a^b f(x)\,dx$. A bonus question asked for a function f and a number a such that $\int_a^x f(t)\,dt = \tan x - 1$. The next year's first hour exam asked whether $\lim_{x \to 0} x^3 \cos(1/x)$ exists, and for justification of the answer. The 1991 final examination provided grids on which to sketch the graphs of a function f and its derivative given information about the signs of f' and f'', limiting information for f, and some function values of f and f'.

3. COMMON FINAL EXAMINATION PERFORMANCE. Every semester from Fall, 1989 through Spring, 1994, all first-year calculus students at the University's main campus—whether in experimental or traditional sections—took a common final examination. There was also a common text for all sections: **[12]** for the first three years and then [**13**]. With the exception of Fall, 1991, when an instructor of a traditional section prepared the common final examination alone, members of the calculus-reform team who taught experimental sections participated with the other instructors in making up the final examination. There were *never* computer-specific questions, but there were always both conceptual questions like those in Section 2 and traditional computational and procedural items. In an effort to measure any possible trade off between conceptual understanding and computational skills, the second-semester examinations tended to put greater emphasis on the latter to balance the heavier conceptual flavor of the first-semester examinations. Table 1

TABLE 1

Term	Standard Section Mean	Experimental Section Mean	Overall Standard Deviation
1989 Fall	59	68	21
1990 Spring	53	56	18
1990 Fall	51	59	18
1991 Spring	59	69	16
1991 Fall	68	71	17
1993 Fall	57	62	18
1994 Spring	62	63	17

presents the data from the results of the common examinations. Although the *size* of the difference is somewhat variable, for every semester the mean score in the experimental sections was higher than that in the standard sections. This suggests that better conceptual understanding in the project sections did not seem to come at the expense of weaker hand-computational skills.

For 1989–90, part of the analysis for a longitudinal study (see Section 4) showed no significant difference between students in the experimental and traditional sections in terms of such success predictors as high-school rank and SAT scores. Such analysis was not carried out for intervening years, but in [**5**] Mary Ann Connors of the University of Massachusetts reports on a comprehensive analysis of the final examination data from 1993–94. That was the first year in which the computer-integrated calculus sequence existed as a separate course (instead of experimental sections within a single first-year calculus sequence). That Fall the two courses were of nearly equal size, a very different distribution from that in Fall, 1989, when 90% of the students were in traditional sections. Enrollment in the two versions was not random: students were free to elect either, and, although

rare, switching between them was possible. Most enrollees were entering freshmen, and despite distribution of full descriptions of the two variants of the course, at the first meetings most students were not aware of any format differences. Analysis of SAT Mathematics and Verbal scores again showed no significant differences between the two populations. Males entered both courses with significantly higher SAT math scores than the females.

Connors determined that the difference in final-examination performance was significant ($p < 0.034$), and that the performance of both males and females from the computer-integrated course was superior to that of their counterparts from the traditional version. Her analysis further revealed that in the computer-integrated course the mean score for female students (61%) was almost identical to the male-student mean (61.4%). Moreover, those female students had a mean SAT Math score (576) lower than both the females (596) and the males (601) in the traditional course. Yet on the final examination they outperformed both those groups, whose respective mean scores were 56.6 and 57.6. As Connors observed, this suggests that the female students benefited more from the computer-integrated experience than did the males. Analysis of performance within the respective courses further supports that. Male students in the computer-integrated course, for instance, had a mean SAT mathematics score of 625, but a mean final examination score just four-tenths of one percent higher than the female mean.

4. CALCULUS AS A PUMP TO TECHNICAL MAJORS. Consistently higher performance of students from reformed sections on concept-rich common final examinations suggested meaningful short-term benefit from the new approach, but to gauge the longer-term effect of the new instructional mode the authors focused on a basic theme of *Calculus for a New Century*. The welcoming message [**21**] from Dr. Frank Press, President of the National Academy of Sciences, to *Calculus for a New Century* cited the role of calculus as a giant filter that knocked many students, especially minorities and women, out of the pipeline to the technical work force of the 21st century. In what was to become the rallying cry of calculus reform, Dr. Robert White, President of the National Academy of Engineering [**28**] followed those remarks with a call to transform calculus from a filter into a pump.

To measure persistence and success in majors in the mathematical, physical, and engineering sciences, the authors first surveyed the requirements of all the science and engineering major programs at the University. From that study they compiled a set of 32 key courses with first-year calculus as prerequisite. Those 32 courses are essential to completion of majors in the fields of interest, and they require students to use the quantitative, analytical, and problem-solving skills that calculus aims to impart. The post-calculus careers of all students who enrolled in first-semester calculus in the Fall of 1989 ($n = 579$) were tracked for the subsequent five academic years. The population was divided into two cohorts for the purposes of this analysis: those who took at least one semester of computer-integrated calculus and those who took two semesters of traditional calculus during 1989–90. The data provided a list of the key courses in which each student subsequently enrolled. Among all students who took first-year calculus in 1989–90, approximately 59% of males took at least one such course, and approximately 43% of females did so. A detailed analysis of each student's performance in all such later courses was performed. A general linear model was fitted to the number of subsequent courses, and the results were checked for consistency with other methods (Mann-Whitney-Wilcoxon and Savage) and by means of transformations such as square-root and logarithmic.

With Mathematics SAT score, enrollment in the project's sections of calculus was one of only two factors that correlated significantly with persistence in technical majors among all students and among male students. For female students, it was the *only* statistically significant factor the study found to correlate with such persistence.

SAT scores, high-school class ranks, and other predictors of success in technical majors were compared both for students who took one or two semesters of the experimental version of calculus and for those who took two semesters in the traditional format. There were no significant differences in SAT scores or class rank, although in both categories the mean was *slightly* higher for students in the experimental sections. However, there was a significant difference (t-test $p < 0.0162$) between the two cohorts in the number of key post-calculus major courses completed. Among students in the traditional sections, the mean number of such courses was 3.50. Among students in the computer-integrated sections, it was 4.95, that is, more than 40% higher. Under the general linear models procedure with SAT scores and high-school rank as covariates, the p-value was 0.0580.

The mean grade earned in those courses was *not* significantly different between the two groups, indicating that students from the traditional sections who persisted in technical majors seem to have acquired adequate preparation from their first-year calculus courses.

Further analysis revealed additional interesting information. That study considered several factors, including SAT Verbal score, SAT Math score, total SAT score, high-school rank, and various combinations of two or more of those in relation to post-calculus persistence in technical fields. For all students, the SAT Math score was a very significant predictor of persistence in such majors ($p < 0.0001$). However, restricted to females it was not significant. In fact, the *only* significant such factor ($p < 0.0267$) for them that this study identified was enrollment in an experimental section of first-year calculus. By contrast, among males only the SAT Math score was a significant ($p < .0024$) factor in persistence.

5. RESULTS FROM OTHER SITES. Encouraging as the foregoing results at Connecticut may be, in themselves they provide information about the experience of just a single institution. The authors investigated the question of how common the Connecticut experience might be at calculus-reform sites. They solicited data from many sources and examined the literature for reports of studies that addressed such areas as persistence in technical fields, performance on common examinations and in later courses, and similarities and differences in outcomes between males and females.

The institution with the most similar approach to Connecticut's is Dartmouth, where True BASIC originated. While the main thrust of [**1**] is a description of Dartmouth's approach and presentation of sample materials, it does mention an experiment in 1988, a year before Connecticut's longitudinal study started. Dartmouth gave precisely the same *traditional* final examination in its first-term introductory calculus course as it had the previous year, when the degree of computer use in the course was substantially lower. It found no differences in performance between the two classes, which led to the Hippocratic conclusion that at least its computer integration appeared to have done no harm.

At the United States Naval Academy, a study compared results of common final examinations for sections that used the calculus-reform approach of the Harvard

Consortium to those for sections that used a traditional text and approach [**19**]. Students were randomly assigned to the sections, with no switching possible. The final examination included a common block of ten multiple-choice equations of both conceptual and computational types. Students from the reform sections scored better overall on nine of those ten questions, with statistically significant disparities in six of them. There was little or no underperformance on traditional computational questions, but significantly better performance on conceptual questions. These results are consistent with the conclusion that the reform mode could foster better learning, especially of conceptual areas of calculus.

Baylor University, another reform site that adopted the Harvard Consortium approach, also appraised the effect by using a common final examination for reform and traditional students [**25**]. Its analysis incorporated an index that divided students by success predictors (SAT Math, ACT Math, and local calculus placement exam scores). Even though *all* common exam questions came from the traditional course's topical coverage and assignments, mean scores among all levels of students in reform sections surpassed those of the traditional sections. On only two of 20 items was the mean score higher in traditional sections.

A study at the United States Merchant Marine Academy analyzed grades of students in traditional and reform versions of calculus throughout the sequence during the years 1990–91 (traditional) and 1991–92 (reform) [**22**]. Overall, the reform-course students earned consistently higher grades than did their predecessors in the traditional version. For students with weaker SAT and calculus-readiness examination scores, that improvement was striking. By contrast, students with the best readiness-exam scores actually did slightly worse in reform calculus, perhaps reflecting the lessened prominence of algebraic manipulation in the latter.

The C^4L (Calculus, Concepts, Computers, Cooperative Learning) program at Purdue has used a two-pronged approach to evaluate its impact. Besides qualitative research on student learning, the C^4L project conducted a longitudinal study similar to Connecticut's over the period 1988–91 [**23**]. Among the variables it studied were the number of post-calculus mathematics courses students took and the grades they earned. Like Connecticut's study, it found that project students took more of those courses, with no significant differences in grade performance.

The University of Illinois at Chicago compared grade performance in subsequent technical courses among students who took a traditional version of calculus in 1994–95 to that of students who took a reformed version of calculus (again, the Harvard approach) in 1995–96. Results of that study showed that students from reformed calculus performed significantly better in physics courses taken immediately after calculus, with diminishing differences in later courses [**2**].

SUNY Stony Brook compared percents of first-semester students who continued into the second semester of calculus before and after adoption of a reformed calculus program. Results from the study show significant differences over two three-year periods (1988–91 and 1992–95) [**20**]. In the first three years, the "yield" (percent of first-semester students who continued to the second semester) was, respectively, 36.4%, 52.6%, and 52.5%. Over the next three years, those figures improved to 63.1%, 59.8%, and 63.4%. One caveat: markedly higher course grades in the reformed course could be a significant stimulant to the improvement, and it is difficult to compare grading standards in the two quite different settings.

Michigan's reform project (Harvard materials) looked at a similar retention pattern. Results showed significant increases the first two years (when project section grades were also significantly higher), but a reverse in the third year when that grade disparity disappeared [**4**]. Significant attitudinal differences on the part

of the students in the reformed version were also identifiable. Those in the reform project's sections showed more positive attitudes about calculus generally and their course experience in particular.

The Duke project (Project CALC) carried out one of the most comprehensive assessment studies, which measured attitude, subsequent enrollment patterns, retention of knowledge, and other factors. No short summary here can do justice to the copious data in that project's final evaluation [**3**]. Suffice it to note that results of the retention tests show a nearly uniform significant advantage in favor of the project students. Their attitudes, problem-solving, and conceptual understanding consistently surpassed those of traditional-section students. However, performance on computational skills was inferior, by a margin that fell just short of statistical significance. Project students again took significantly more technical courses, but students from traditional sections had a slightly higher grade-point average in those courses. Significantly more project students than traditional ones took more than two post-calculus courses.

What about the gender-specific aspects of the Connecticut project? The positive results for females in the Connecticut study are consistent with reports of similar effects in projects that incorporate graphing-calculator technology; see [**7**] and [**18**]. An interesting question that comes from such data: what aspects of technology enhance female learning in quantitative subjects? One possible theory from informal discussions is the prominent role of cooperative learning in many calculus-reform projects. Computer-laboratory sessions and related projects lend themselves naturally to group work, which several studies (among them [**11**]) have correlated with improved performance by females in science and mathematics courses.

Finally, an earlier article in this MONTHLY reported about calculus reform at Oklahoma State University [**17**]. That study found that—as at the Merchant Marine Academy, SUNY Stony Brook, and Michigan—grades in the reform version of calculus were higher than in traditional sections, which in the Oklahoma State study were contemporaneous. Little difference in subsequent enrollment patterns *within the calculus/differential equations sequence* was observed. However, the study did not encompass a sufficiently long time span to measure persistence in and completion of required courses for technical majors, as in the Connecticut, Purdue, and Duke studies. However, its results were consistent with those from the Merchant Marine Academy study in one respect: lower grades in traditional sections of Calculus II for students from reform sections of Calculus I. It reported insufficient data for higher-level course performance to permit meaningful conclusions, but stated that fewer students from the reform sections earned C or better grades in differential equations or linear algebra. It concluded that traditional-section students tended overall to do better in subsequent mathematics courses, but gave no information about the statistical significance of their performance.

The preponderance of evidence from the Connecticut study—as well as the evidence cited in this section—is consistent with the conclusion that the impact of calculus reform has been positive. If the "filter-to-pump" goal is an appropriate assessment standard, then the results from Connecticut and other calculus-reform sites suggest that calculus with modern computational and pedagogical features can promote better end-of-course mastery, significantly improve the flow rate into the technical work force, and foster more gender diversity in that flow. Note the caution in these conclusions: *is consistent with* and *suggest*, rather than *imply* or *prove*. Unlike mathematics, appraisal of pedagogical change is a highly inexact science, lacking not only proofs but even generally accepted rules of inference.

The positive content of the previous paragraph notwithstanding, it is only fair to mention that at many institutions the fruits of calculus reform have not included substantially higher performance and persistence [**9**]. Experimenters whose results fall short of hopes and expectations are certainly less likely to write up and disseminate accounts of their work than those with more pleasing performance data to report. There is also considerable belief in the *Hawthorne effect*, that is, that expenditure of time and attention on a project aiming to improve outcomes leads to actual improvements, which to at least some extent result from the attention itself rather than the project's methods. On the other hand, the atmosphere of passionate debate mentioned in Section 1 certainly is receptive to a well-designed and well-documented study showing predominantly *lower* performance and/or persistence by students in a reformed approach to calculus. The authors are not aware of any such study.

6. GENERAL NATIONAL IMPACT. On a procedural level, there is clear evidence of substantial change in the teaching, learning, and testing of calculus since the 1980's. The appearance and spread of graphing calculators and powerful desktop, laptop, and now notebook and subnotebook computers has noticeably stimulated greater emphasis on the graphical and numerical aspects of calculus. The popularity of projects and written reports has also brought a new verbal dimension to the subject. Even current editions of texts that early reformers cited as contributors to the calculus crisis incorporate (and advertise) numerical, graphical, and verbal components. Calculator and computer supplements, which were virtually nonexistent at the start of the calculus-reform movement, have now become pervasive. It is in fact challenging to identify a current calculus text that in 1987 would have been labeled "traditional."

National examinations reflect national norms about mastery of the subjects they test. Prior to the calculus reform movement, the Advanced Placement calculus program banned calculators. The current AP exams *require* them! A similar situation will shortly exist in the new Graduate Record Examination's Mathematical Reasoning Test for students planning to pursue graduate study in engineering, the physical, mathematical, or computer sciences, economics, or some areas of the life sciences [**26**]. This test presumes a year of calculus, and the questions in its new version probe understanding of concepts and their reflection in graphical or applied settings, areas of particular prominence in calculus reform.

The foregoing sections raise a natural question: what national impact has calculus reform actually had? The rest of this section addresses that question.

More than 500 mathematics departments at postsecondary institutions nationwide are currently implementing some level of calculus reform. These "reformed" courses enroll approximately 300,000 students each year, about 32% of the total national calculus enrollment [**27**]. Such growth of a movement only a decade old suggests that the influence of calculus reform will continue to spread. In fact, many institutions are now initiating programs to improve learning in a wide range of science, mathematics, engineering, and technology courses. These include projects that eliminate the traditional boundaries between these disciplines by means of an integrated teaching approach. Such programs, which respond to the new requirements students face in an increasingly technical and multidimensional workplace, represent a fundamental change in the philosophy that has long guided the structure of undergraduate education.

At present, there are few studies that document the impact of these efforts on student learning, faculty and student attitudes, and the general environment at

undergraduate institutions. There has been considerable study of student learning in calculus, e.g. [**6**], [**8**], [**9**], [**10**], and [**24**]. To make meaningful judgments about the value of reform efforts, it is necessary to understand not only how students learn, but also how different environments influence their ability to learn. More studies are essential if the academic community is to understand the implications of this change in philosophy on learning within and across disciplines, throughout a student's experience at the undergraduate level, and beyond.

One such study is part of a larger effort by the National Science Foundation to evaluate the impact of reform in science, mathematics, engineering, and technology education at the undergraduate level [**9**]. It investigated the effect of calculus reform on: student achievement, attitudes, and retention; the implementation of mechanisms that have shown promise in improving the learning environment (e.g., student-centered learning, multiple methods of delivery, and alternative measures of assessment); and the general educational environment. Although the study yielded mixed results for student achievement, several trends did emerge. For example, approximately 50% of the institutions conducting studies on the impact of technology reported increases in conceptual understanding, greater facility with visualization and graphical representations, and the ability to solve a wider variety of more difficult problems, without any loss of computational skills. Another 40% reported that students in classes with technology had done at least as well as those in traditional courses. The impact of long-term projects and group work on student achievement is less clear. However, a pattern was discernible regarding the type of student likely to excel in this environment. This includes students who are "above average" in mathematics, students who do not perform well on traditional tests, and engineering majors.

Projects and group work also seem to affect grade distribution, although not in consistent ways. For example, one institution reported that projects made the final grade distribution more bi-polar, with very few "C" students, while another reported that projects were "the great equalizer," with more C's resulting from the subjective nature of grading the projects and concomitant inability to justify grades at the extremes. The effect of projects on course grades also seemed to be influenced by whether the projects were individual or group: the latter seemed to "equalize" the grade distribution more than individual assignments.

The existence of many common elements throughout the majority of the projects implies that the relative success or failure of reform efforts is not necessarily dependent solely upon content, but also on how, by whom, and in what setting the approach is implemented. Consistent reactions of students from a wide variety of institutions point to several key components for a successful reform environment. For example, instructors must communicate to students (and other faculty) the purpose of the changes being made in the calculus course. This is often not as easy as it may seem, because the reasons for the change must strike the students as relevant and important to future success. Unfortunately, many students believe mathematics is a static list of rules and algorithms to be memorized, a barrier to be overcome before they can do "real" work in other disciplines. One of the most important roles of reform efforts, then, is to challenge those beliefs and help students appreciate the many uses of calculus, both within mathematics and in a wide range of other disciplines. Thus far, the most effective means to this end still await identification.

Two major goals of calculus reform have been to revitalize the sequence and to generate discussion within the mathematics community about the nature and content of the calculus course. The lively debate the introduction cites attests to

the fact that these goals have indeed been realized! In itself, that constitutes a major accomplishment that should not be overlooked. Are the reform efforts in fact helping students better understand calculus and appreciate the importance of mathematics in their lives? This basic question may never be definitively answered, but it cries out for study if calculus courses—and mathematics in general—are to remain a vital part of the undergraduate curriculum.

ACKNOWLEDGMENTS. An earlier version of this article was presented under the title *The impact of computer-integration in a first-year calculus sequence* to the AMS/MAA Special Session on Research in Undergraduate Mathematics Education at San Diego, January 9, 1997.

The authors take pleasure in thanking their colleagues from other reform sites, who generously responded to requests for data and results from their assessment work. In many cases they furnished otherwise unpublished information or internal institutional reports. The referees also provided many helpful suggestions, including sources of assessment data. The first author is most grateful to the National Science Foundation for its support (Grants DMS-8852395, USE-9153270, DUE-9252463), as well as to the Elias Howe High-Technology Program and to his administration for continuing support of calculus reform. Finally, he thanks his colleagues and students for their good cheer and encouragement throughout the project.

REFERENCES

1. J. Baumgartner and T. Shemanske, Teaching calculus with True BASIC, *Priming the Calculus Pump: Innovations and Resources*, MAA Notes **17**, Washington, DC, Math. Assn. of America, 1990, pp. 33–50.
2. J. Baxter, D. Majumdar, and S. Smith, Subsequent-grades assessment of traditional and reform calculus, *PRIMUS* **8** (1998) 317–330.
3. J. Bookman and C. Friedman, Evaluation of Project CALC 1989–1993, report to National Science Foundation, 1994.
4. M. Brown, The new calculus program at the University of Michigan, Univ. of Michigan Dept. of Math. reports, 1993–1996.
5. M. Connors, Achievement and gender in computer-integrated calculus, *J. Women and Minorities in Sci. Engr.* **2** (1995) 113–121.
6. M. Davidson, *Cooperative Learning in Mathematics: A Handbook for Teachers*, Menlo Park, CA, Addison-Wesley, 1990.
7. P. Dunham, Calculator and gender issues, *Assn. of Women in Math. Newsletter* **25** (2) (1995) 16–18.
8. S. Ganter, The impact of ten years of calculus reform on student learning and attitudes, *Assoc. for Women in Sci. Magazine* **26** (1997) 10–15.
9. S. Ganter, Ten Years of Calculus Reform: A report on evaluation efforts and national impact, MAA Notes, to appear.
10. M. Heid, Resequencing skills and concepts in applied calculus using the computer as a tool, *J. Res. in Math. Educ.* **19** (1988) 3–25.
11. C. Hurley, Cooperative learning in the quiz section of general chemistry, *Proc. Frontiers in Educ.*, Washington, DC, IEEE & ASEE, 1993, pp. 162–166.
12. J. Hurley, *Calculus*, Belmont, CA, Wadsworth Publ. Co., 1987.
13. J. Hurley, *Calculus: A Contemporary Approach*, New York, McGraw-Hill, Inc., 1992.
14. J. Hurley, Computer-integrated calculus at the University of Connecticut, *Proc. Fifth Intl. Conf. Technology in Collegiate Math.*, Addison-Wesley, 1994, pp. 777–784.
15. J. Hurley, C. Vinsonhaler, and W. Wickless, *Student Manual for Computer-Integrated Calculus I, II*, New York, McGraw-Hill, Inc., 1995 (Vol. I), 1996 (Vol. II).
16. J. Hurley, C. Vinsonhaler, and W. Wickless, *Instructor Guide for Computer-Integrated Calculus I, II*, Storrs, CT, Univ. of Connecticut Math. Dept., 1995 (Vol. I), 1996 (Vol. II).
17. K. Johnson, Harvard calculus at Oklahoma State University, this MONTHLY **102** (1995) 794–797.
18. P. Jones and M. Boers, Some gender differences in attitudes and mathematics performance with graphics calculators, *Proc. Fifth Intl. Conf. Technology in Collegiate Math.*, Addison-Wesley, 1994, pp. 173–177.
19. H. L. Penn, Comparison of test scores in Calculus I at the Naval Academy, *Focus on Calculus* (Wiley) **6** (1994) 6–7.

20. A. Phillips, Persistence MAT 131 → MAT 132, SUNY Stony Brook Dept. of Math. Report, `http://www.math.sunysb.edu/~tony/calc/finrep/finrep.html`, 1995.
21. F. Press, Calculus for a new century, *Calculus for a New Century*, MAA Notes **8** (1987), Washington, DC, Math. Assn. of America, p. 5.
22. G. Ratay, Student performance with calculus reform at the United States Merchant Marine Academy, *Primus* **3** (1993) 107–111.
23. K. E. Schwingendorf, Assessing the Effectiveness of Innovative Educational Reform Efforts, in B. Gold, S. Keith, W. Marion, eds., *Assessment Practices in Undergraduate Mathematics*, MAA Notes Series 49, Mathematical Association of America, Washington, D.C., 1999, pp. 249–252.
24. D. Tall, Inconsistencies in the learning of calculus and analysis, *Focus on Learning Problems in Mathematics* **12** (1990), Center for Teaching/Learning of Math., Framingham, MA, pp. 49–63.
25. E. Tidmore, A comparison of calculus materials used at Baylor University, *Focus on Calculus* (Wiley) **7** (1994) 5–6.
26. A. Tucker, New GRE mathematical reasoning test, *Notices Amer. Math. Soc.* **42** (1995) 245–247.
27. A. Tucker and J. Leitzel, Eds., *Assessing Calculus Reform Efforts: A Report to the Community*, MAA Reports **6** (1994), Washington, DC, Math. Assn. of America.
28. R. White, Calculus of reality, *Calculus for a New Century*, MAA Notes **8** (1987), Washington, DC, Math. Assn. of America, pp. 6–9.
29. H. Wu, The mathematics education reform: why you should be concerned and what you can do, this MONTHLY **104** (1997) 946–954.

JIM HURLEY received his Ph.D at UCLA under the direction of Robert Steinberg. Since 1971 he has been at the University of Connecticut, which named him a University Teaching Fellow in 1995. Many of his instructional innovations are stolen from his wife, Cecile, who directs UConn's general chemistry program. His research interests center on the structure theory of Lie algebras and related groups. Avocations include gardening, philately, and basketball—recently, again following his spouse's lead, cheering on the UConn women's team.
University of Connecticut, U-9, Storrs CT 06269-3009
hurley@math.uconn.edu

UWE KOEHN received his BS from Queens College (CUNY) and Ph.D from the University of Illinois at Champaign-Urbana. He recently took Professor Emeritus status after 30 years in the Statistics Department at the University of Connecticut—12 of those years as Department Head. To the extent that it does not interfere with his opera subscription, he does research and statistical consulting, and teaches a one-credit course of elementary statistics, in German.
Department of Statistics, University of Connecticut, U-120, Storrs CT 06269-3120
qqqq@uconnvm.uconn.edu

SUSAN GANTER completed graduate work in applied mathematics, evaluation, mathematics education, and administration at UC, Santa Barbara. Since her Ph.D she has held faculty mathematics positions at Western Washington University, CSU San Bernardino, and Worcester Polytechnic Institute. In 1996, she received a Senior Research Fellowship from the American Educational Research Association and the National Science Foundation; it supported work to develop and evaluate national programs for the improvement of undergraduate science, mathematics, engineering, and technology education. After leaving NSF in 1998, she continued those efforts at the American Association for Higher Education until this fall, when she joined the Mathematical Sciences faculty at Clemson University.
Department of Mathematical Sciences, Clemson University, Clemson, SC 29634-1907
sganter@clemson.edu

Some Fundamental Control Theory II: Feedback Linearization of Single Input Nonlinear Systems

William J. Terrell

1. INTRODUCTION. In Part I of this article [**12**] we characterized the single-input single-output linear

$$x' = Ax + bu(t) \tag{1a}$$

$$y = c^T x \tag{1b}$$

where A is $n \times n$, x and b are $n \times 1$, and c is $n \times 1$, that can be transformed by a nonsingular linear transformation, $z = Tx$, to a linear *companion form*

$$z' = Pz + du(t) \equiv \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & 1 \\ -k_n & -k_{n-1} & -k_{n-2} & \cdots & -k_1 \end{bmatrix} z + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} u, \tag{2}$$

which has the desirable property of *complete controllability*. By the use of state feedback, $u = Kx$ with K a $1 \times n$ matrix, such systems may be expressed in the particularly simple form, $z_1^{(n)} = v$, where v is a new reference input that is available for control purposes.

For convenience we summarize the main result of Part I in Theorem 1; see [**12**] for definitions of the relevant concepts.

Theorem 1. *System* (1a) *can be transformed by a nonsingular linear transformation, $z = Tx$, to the companion system* (2), *if and only if rank* $[b\, Ab \ldots A^{n-1}b] = n$ ((1a) *is completely controllable*). *When this is the case, T is unique and*

$$T = \begin{bmatrix} \tau \\ \tau A \\ \vdots \\ \tau A^{n-1} \end{bmatrix}, \tag{3}$$

where τ is the unique solution of

$$\tau\, [b\, Ab \ldots A^{n-1}b] = [0 \ldots 0\ 1] = d^T. \tag{4}$$

There exists a nonsingular transformation $z = Tx$ taking (1a) *to the companion system* (2) *with $z_1 = y = c^Tx$, if and only if*

$$\mathit{rank} \begin{bmatrix} c^T \\ c^TA \\ \vdots \\ c^TA^{n-1} \end{bmatrix} = n,$$

(*system* (1a), (1b) *is completely observable*) *and*

$$c^T[b\,Ab\ldots A^{n-1}b] = [0\ldots 0\;1] = d^T.$$

When this is the case, T *is unique and is the matrix* (3) *with* $\tau = c^T$.

Our main goal in Part II is to generalize Theorem 1 for the simplest form of single input nonlinear systems. The exposition uses ideas from the theory of differential equations, linear algebra, and analysis, and basic concepts in differential geometry. The equivalence problem we consider is one of the fundamental results of geometric nonlinear control theory. The original formulation and solution of this equivalence problem in the single-input case is due to R. W. Brockett [**1**].

We begin with a calculation that recapitulates the essential calculations that established Theorem 1. It is also useful in the generalization of Theorem 1. Consider the following product of $n \times n$ matrices,

$$\begin{bmatrix} c^T \\ c^TA \\ \vdots \\ c^TA^{n-1} \end{bmatrix}[b\,Ab\,\cdots\,A^{n-1}b] = \begin{bmatrix} 0 & \cdots & \cdots & 0 & c^TA^{n-1}b \\ 0 & & & * & * \\ \vdots & & & & \vdots \\ 0 & * & & & \vdots \\ c^TA^{n-1}b & * & \cdots & \cdots & * \end{bmatrix}, \quad (5)$$

which results when c, A, b satisfy $c^TA^kb = 0$ for $0 \le k \le n-2$. If $c^TA^{n-1}b \neq 0$ as well, then both matrices on the left must be nonsingular. In Part I we showed that, conversely, if one of the matrices on the left is nonsingular (for *some* b, respectively c), then the other matrix is *also* nonsingular (for some c, respectively b). Nonsingularity on the right in (5) implies an observability condition and a controllability condition [**12**]. The geometric interpretation of the zeros on the right hand side is that the null space of the linear functional $y = c^Tx$ is the $(n-1)$ dimensional space, *span*$\{b, Ab, \ldots, A^{n-2}b\}$. The duality aspects of Part I arise from the duality between vectors and linear functionals in a finite dimensional vector space. Similar duality considerations in Part II involve the pairing of *vector fields* and *co-vector fields* (or *differential* 1-*forms*).

If we replace the linear vector field Ax in (1) by $f(x)$, and the vector b by a vector field $g(x)$, we are interested in determining exactly when the resulting single input nonlinear system $x' = f(x) + g(x)u$ may be transformed to the special form $y^{(n)} = v$ *by local coordinate change and state feedback*. If such a transformation is possible, one can exploit the special control-theoretic properties discussed in [**12**] for that special form. In such a case, a nonlinear system can be controlled using linear control methods.

Example 1. Suppose $x \in R^3$ and $u \in R$. Is the system

$$x' = f(x) + g(x)u \equiv \begin{bmatrix} \exp(-x_2) \\ x_1 \\ \frac{1}{2}x_1^2 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} u$$

equivalent to $y^{(3)} = v$? What is this equivalence, and how do we determine if such an equivalence is possible?

2. LINEARIZATION AND FUNCTIONS OF RELATIVE DEGREE n. Consider the single-input single-output nonlinear version of (1a) described by

$$x' = f(x) + g(x)u, \quad (6a)$$

where f and g are smooth vector fields defined in some open region $\mathscr{G}$ of R^n. The term *smooth* means that the components of f and g are continuously differentiable as often as required in our discussion. Equation (6a) may be augmented with an output,

$$y = h(x), \tag{6b}$$

where h is a smooth real-valued function defined on $\mathscr{G}$. We are interested in necessary and sufficient conditions under which system (6a) is equivalent to a linear companion system. Such an equivalence, if possible, generally requires feedback in addition to coordinate change: our examples make it clear that (4), the condition that allows for a coordinate transformation *directly* to companion form in the linear case, cannot generally be expected here.

2.1 Input-Output Linearization. As in the linear case, we can try to use a known output h to help define a local coordinate change, $z = T(x)$, and state feedback, $u = \alpha(x) + \beta(x)v$ (with $\beta(x) \neq 0$), in a neighborhood U of x_0 in (6), to produce a linear input-output equation,

$$y^{(j)} = v. \tag{7}$$

In (7), v is called the new *reference input*, and we would like to have $j = n$ in order to capture the full n-dimensional dynamical system. This is the Input-Output Linearization Problem (IOLP). The idea is that feedback by $u = \alpha(x)$ simplifies the system equations by cancelling nonlinearities when possible, and then subsequent controls v are available to control the dynamics in a desired manner.

The next definition recalls the behavior of the special outputs that yield observability *and* a companion form in the linear case of [**12**]. The definition is motivated by the form of the terms that appear when a function is differentiated repeatedly along system (6a).

Definition 1. The Lie derivative of a real-valued function h along a vector field g is the real-valued function $L_g h$ defined by $L_g h(x) \equiv dh(x) \cdot g(x)$, where $dh(x)$ is the row gradient of h at x. For iterated derivatives of this type we write $L_g^0 h = h$, and $L_g^k h = L_g(L_g^{k-1} h)$. The function h has *relative degree* $j \geq 1$ with respect to (6a) at the point x_0 if

(i) $L_g L_f^k h(x) = 0$ for all x in a neighborhood of x_0, for all $0 \leq k < j - 1$, and
(ii) $L_g L_f^{j-1} h(x_0) \neq 0$.

The relative degree is the number of differentiations of h along the system that are needed to make u appear explicitly. Thus, for relative degree $j = 1$, note that only condition (ii) is relevant. There may be points x_0 where a relative degree for a function is not defined. By continuity, conditions (i) and (ii) allow us to speak of a function h having relative degree j *in an open set* U containing x_0.

Example 2. Consider the system

$$x_1' = \sin x_2 \tag{8}$$

$$x_2' = -x_1^2 + u. \tag{9}$$

The function $y = h(x) = x_2$ has relative degree 1 at all points since $L_g h(x) = 1$. On the other hand, $y = h(x) = x_1$ satisfies $L_g h = 0$, and $L_g L_f h(x) = L_g(\sin x_2) = \cos x_2$, so this function has relative degree 2 at the equilibrium $x_0 = 0$ of the unforced system.

Functions having relative degree n are especially useful. Theorem 1 illustrates why linear functionals of relative degree n were central in [**12**].

Suppose an output $y = h(x)$ has relative degree n at x_0. Then y and the first $n - 1$ time derivatives of y along (6a) yield the equations

$$\begin{bmatrix} y \\ y' \\ \vdots \\ y^{(n-1)} \end{bmatrix} = \begin{bmatrix} h(x) \\ L_f h(x) \\ \vdots \\ L_f^{n-1} h(x) \end{bmatrix} + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ L_g L_f^{n-1} h(x) \end{bmatrix} u. \tag{10}$$

In Proposition 2 we show that if h has relative degree n at x_0, then the vector function $[h \ L_f h \ \cdots \ L_f^{n-1} h]^T$ has nonsingular Jacobian with respect to x at x_0; thus, the functions $L_f^k h(x)$ for $0 \le k \le n-1$ can be used as component functions in a nonlinear coordinate transformation defined locally around x_0. The presence of the nonzero function $L_g L_f^{n-1} h(x)$ as a coefficient of u prevents us from obtaining the linear, controllable companion form for $z = [y, y', \dots, y^{(n-1)}]$ by coordinate transformation alone. However, the coordinate transformation *together with state feedback* does produce the simple form, $y^{(n)} = v$. We need only define

$$u = \frac{1}{L_g L_f^{n-1} h(x)} \left(-L_f^{n-1} h(x) + v \right), \tag{11}$$

where v is the new reference input. This situation allows for control action on the nonlinear system by operating with the linear controllable form (7) (with $j = n$). Note that an assumption of relative degree $j < n$ leads to a linear input-output relation (7). Relative degree n of h at x_0 is a type of local observability condition: with $u \equiv 0$, the Inverse Function Theorem [**9**, p. 193] applied to (10) implies that x is determined by $[y, y', \dots, y^{(n-1)}]$ near x_0.

2.2 Input-State Linearization. If a relative degree n function is not readily available as an output function, we still ask for a local coordinate change, $z = T(x)$, and feedback $u = \alpha(x) + \beta(x)v$ with $\beta(x) \neq 0$ near x_0, that produces a *controllable linear system* for z. This is often called the Input-State Linearization Problem (ISLP). As shown in [**12**], given a controllable linear system, an additional coordinate change and state feedback may be used to produce the linear system

$$z' = Nz + dv \equiv \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} z + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} u, \tag{12}$$

where N is the standard nilpotent block with ones on the superdiagonal and zeros elsewhere, and $d = [0 \dots 0 \ 1]^T$ as usual. Thus we take the ISLP to mean the problem of achieving the form (12).

Our first goal is to show that the ISLP is solvable if and only if there exists a function of relative degree n at x_0 for (6a). We have already discussed the sufficiency of the relative degree n condition, though the proof of Proposition 2 is needed to complete that discussion. After the next example, we show that solvability of the ISLP implies the existence of a function $\lambda(x)$ of relative degree n at x_0 —namely, the first coordinate function T_1 of the transformation T. Our ultimate goal is to derive computable geometric conditions for the existence of a relative degree n function.

One can adopt the point of view that there is no essential mathematical difference between the linearization problems IOLP and ISLP when a relative degree n function exists—the difference lies in whether there is an *explicitly known* output function of relative degree n, or if such a function must be *constructed.* From the practical point of view, of course, this is an essential difference. We return to Example 2 to illustrate an important consideration.

Example 3. Consider again the system,

$$x_1' = \sin x_2$$

$$x_2' = -x_1^2 + u.$$

It is easy to linearize the input-output behavior; take $y = x_2$, and set $u = x_1^2 + v$, which produces the linear input-output relation $y' = v$. This relation would make it easy to track a *specified* output $y(t) = x_2(t)$ by control v, but the linearizing control u would make x_1 unobservable if it is only the resulting input-output relation $y' = v$ that is considered. Thus, some of the original two-dimensional dynamics is hidden by the input-output relation. The presence of such unobservable dynamics introduces the question of internal stability for that dynamics; the x_1 variable may not remain well-behaved (e.g., bounded) when using a control strategy based on the relation $y' = v$. For example, suppose we wanted to hold the output $y = x_2$ via feedback at a constant value, $x_2 = c$. Then the x_1 solution would be $x_1(t) = x_1(0) + t\sin c$, and therefore $x_1 \to \infty$ as $t \to \infty$.

We now show that a solution of the ISLP entails a relative degree n function. Suppose the ISLP is solved, so that the transformation $z = T(x)$, combined with feedback $u = \alpha(x) + \beta(x)v$, produces the linear system (12). From the definitions of the variables, this occurs if and only if

$$\frac{\partial T}{\partial x}(f(x) + g(x)u) = NT(x) + dv, \tag{13}$$

where equality holds *for all* $u = \alpha(x) + \beta(x)v$, with v arbitrary. By considering $u = 0$ and $u = 1$, we see that (13) is equivalent to the two partial differential equations:

$$\frac{\partial T}{\partial x}f(x) = NT(x) - d\alpha(x)\beta^{-1}(x), \tag{14a}$$

$$\frac{\partial T}{\partial x}g(x) = d\beta^{-1}(x). \tag{14b}$$

If $T(x) = [T_1(x)\,T_2(x)\dots T_n(x)]^T$, then

$$NT(x) - d\alpha(x)\beta^{-1}(x) = \left[T_2(x)\,T_3(x)\dots T_n(x) - \alpha(x)\beta^{-1}(x)\right]^T,$$

and

$$d\beta^{-1}(x) = \left[0\,0\dots0\beta^{-1}(x)\right]^T.$$

This allows us to display (14a) in detail:

$$dT_k(x)\cdot f(x) = L_f T_k(x) = T_{k+1}(x), \quad k = 1,\dots,n-1, \tag{15a}$$

and

$$dT_n(x)\cdot f(x) = -\alpha(x)\beta^{-1}(x). \tag{15b}$$

Therefore all components of T are determined by the *first* component T_1. And the last equation specifies the ratio, $-\alpha(x)\beta^{-1}(x)$, in terms of T_1. Similarly we display (14b) using (15a) and Definition 1:

$$dT_{k+1}(x)\cdot g(x) = L_g L_f^k T_1(x) = 0, \quad k = 0,\dots,n-2, \tag{16a}$$

while

$$dT_n(x)\cdot g(x) = L_g L_f^{n-1} T_1(x) = \beta^{-1}(x) \neq 0. \tag{16b}$$

Thus, by (15a) and (16), *the solution of the ISLP requires a function* $T_1(x)$ *that has relative degree* n. Given T_1, the rest of T is then defined by (15a), and the required feedback $u = \alpha(x) + \beta(x)\, v$ is determined by

$$\beta(x) = \frac{1}{L_g L_f^{n-1} T_1(x)}; \quad \alpha(x) = -\frac{L_f^{n-1} T_1(x)}{L_g L_f^{n-1} T_1(x)}. \tag{17}$$

Notice that an equilibrium point of interest, say x_0 such that $f(x_0) = 0$, can always be transformed to $z = T(x_0) = 0$: by (15a), we need only require that $T_1(x_0) = 0$. We summarize the preceding discussion as follows.

Theorem 2. *System* (6a) *can be transformed by coordinate transformation* $z = T(x)$ *and state feedback* $u = \alpha(x) + \beta(x)v$ *in a neighborhood of* x_0 *to the linear controllable form* $z' = Nz + dv$ *if and only if there exists a function* λ *having relative degree* n *with respect to* (6a) *at the point* x_0. *When this is the case,* T *is determined by* (15a) *by defining the first component function by* $T_1 = \lambda$, *and the feedback is determined by* (17).

Proof: Everything has been proved except the legitimacy of the coordinate change

$$z = T(x) = \left[\lambda(x), L_f\lambda(x),\dots, L_f^{n-1}\lambda(x)\right]^T,$$

when λ has relative degree n at x_0. This argument appears as part of a more general statement in Proposition 2. Assuming this result, Theorem 2 follows. ■

Let us illustrate these ideas, and the general *nonuniqueness* of T in the nonlinear case:

Example 4. Consider the system in Example 3 near the equilibrium $x_0 = 0$ of the unforced system (with $u = 0$). The equations (16) for T_1 are

$$L_g T_1(x) = 0, \quad L_g T_2 \neq 0; \quad T_2 = dT_1(x)\cdot f(x).$$

Using $g = [0\ 1]^T$ we see that T_1 is independent of x_2, so $T_2(x) = (\partial T_1/\partial x_1)\sin x_2$. The nontriviality condition becomes

$$L_g T_2 = \frac{\partial T_2}{\partial x_2} = \frac{\partial T_1}{\partial x_1}\cos x_2 \neq 0,$$

which holds as long as $\cos x_2 \neq 0$ and $\partial T_1/\partial x_1 \neq 0$. One solution is $T_1(x) = x_1$, which we used in Example 3. The choice is not unique: $T_1(x) = x_1 - x_1^2$ also works near $x_0 = 0$ in fulfilling the conditions of Theorem 2.

Example 5. Using Theorem 2 we can show that Example 1 is not input-state linearizable. The vector field g in Example 1 is $g = [1\ \ 0\ \ 0]^T$. Let λ be a smooth function, and suppose $L_g\lambda(x) = (\partial\lambda/\partial x_1)(x) = 0$ for x in an open set U; then λ

is independent of x_1 in U, so $\lambda = \lambda(x_2, x_3)$ there. If we also impose the condition that $L_g L_f \lambda(x) = 0$ in U, then we have

$$L_g(d\lambda \cdot f)(x) = \frac{\partial}{\partial x_1}\left(\frac{\partial \lambda}{\partial x_2} x_1 + \frac{\partial \lambda}{\partial x_3}\frac{1}{2}x_1^2\right)$$
$$= \frac{\partial \lambda}{\partial x_2} + x_1 \frac{\partial \lambda}{\partial x_3} = 0;$$

and it is not possible for this to hold in U when $\lambda = \lambda(x_2, x_3)$. Thus, there is no function having relative degree 3 with respect to this system in any open set in R^3.

3. NECESSARY CONDITIONS FOR ISLP SOLVABILITY. We've seen that an "artificial output" T_1 of relative degree n is determined by a *nontrivial* solution of the partial differential equation system (16a), the nontriviality condition being (16b). When is this first order partial differential equation system solvable? We want computable conditions for solvability directly in terms of f and g. Let us first consider necessary conditions, and for that, a nonlinear version of the calculation in (5) is helpful, with an appropriate generalization of the columns b, $Ab, \ldots$ in (5). This analysis of necessary conditions completes the proof of Theorem 2 and leads to the computable criteria we seek. We should remark that there are indeed local observability and controllability conditions involved in what follows, as you might expect. After all, we are transforming to a linear system with the properties of controllability and observability. We discuss a local *reachability* property after the main Theorem 3—this property is related to, but weaker than, complete controllability. For now, we continue to focus on the goal of generalizing the transformation to companion form statements of Theorem 1, but in the process we indicate the intuition involved in generalizations of the linear controllability criterion.

Let us consider (5): we replace the first matrix on the left by $[d\lambda\, dL_f\lambda \ldots dL_f^{n-1}\lambda]^T$, where λ has relative degree n, and we put $g(x)$ in place of b in the second matrix on the left; we then need vector fields to replace $Ab, \ldots, A^{n-1}b$. That is, we need to identify the null space of the differential $d\lambda$. The appropriate vector fields can be motivated either analytically or algebraically, and we consider both aspects in order to build some intuition.

The next definition can be motivated by the calculation of Proposition 1, but it is convenient to state it here and follow it with an important example.

Definition 2. The Lie bracket $[g_1, g_2]$ of two vector fields g_1, g_2 is the vector field defined by

$$[g_1, g_2](x) \equiv \frac{\partial g_2}{\partial x}(x) g_1(x) - \frac{\partial g_1}{\partial x}(x) g_2(x). \tag{18}$$

Here is a notation that helps with iterated brackets: define $ad_{g_1}^0 g_2 = g_2$, $ad_{g_1} g_2 = [g_1, g_2]$, and $ad_{g_1}^k g_2 \equiv [g_1, ad_{g_1}^{k-1} g_2]$ for $k \geq 1$. The brackets described in (18) are important in the linear system case:

Example 6. If $f(x) = Ax$ and $g(x) = b$, then $[f, g](x) = -Ab$. Also, $ad_f^2 g(x) = [f, [f, g]](x) = A^2 b$, and in general, $ad_f^k g(x) = (-1)^k A^k b$.

Example 6 suggests that the brackets $ad_f^k g(x)$ are important in the nonlinear case. To confirm this, we consider an analytic relation involving f and g. Write

$\phi_t^\eta(x)$ for the time t solution map of the differential equation $x' = \eta(x)$; that is, $\phi_0^\eta(x) = x$ and $\partial\phi_t^\eta(x)/\partial t = \eta(\phi_t^\eta(x))$. Then $(\phi_t^\eta)^{-1} = \phi_{-t}^\eta$ where defined. Also write $(\phi_t^\eta)_*$ for the derivative map with Jacobian matrix $\partial\phi_t^\eta/\partial x$. The *variational differential equation* satisfied by $(\phi_t^\eta)_*$ is the equation

$$\frac{\partial}{\partial t}(\phi_t^\eta)_* = \frac{\partial\eta}{\partial x}(\phi_t^\eta)_*;$$

to prove this relation, one differentiates the identity $d/dt(\phi_t^\eta(x)) = \eta(\phi_t^\eta(x))$ with respect to x and then interchanges the order of differentiations with respect to x and t. It is also useful to have the variational equation for $(\phi_{-t}^\eta)_* = (\phi_t^\eta)_*^{-1}$: the formula

$$\frac{d}{dt}A^{-1}(t) = -A^{-1}(t)\frac{dA}{dt}A^{-1}(t)$$

for the derivative of the inverse of a nonsingular matrix function of t, applied to $A = (\phi_t^\eta)_*$, gives

$$\frac{\partial}{\partial t}(\phi_{-t}^\eta)_* = -(\phi_{-t}^\eta)_*\frac{\partial\eta}{\partial x}.$$

Now consider following the flow of f (the vector field in (1a)) for a short time t, to the point $\phi_t^f(x_0)$; then determine the direction vector $(f+g)(\phi_t^f(x_0))$, the direction you would move if you turned the control u "on" with $u = 1$; and, finally, transfer this tangent vector back to the point x_0 by applying $(\phi_{-t}^f)_*$. Thus, consider the "curve of tangents" based at x_0,

$$V(t) = (\phi_{-t}^f)_*\eta(\phi_t^f(x_0)), \tag{19}$$

where $\eta = f + g$, and more specifically, the derivative $V'(0)$ at $t = 0$. By considering shorter and shorter times t, it is plausible that the vector $V'(0)$ (and also the higher order derivatives of V at $t = 0$) should indicate something about the directions we might move, starting at x_0, by some suitable "off-on" switching strategy for the input u. This can be made precise, in a way that provides an alternative motivation for the Lie bracket operation; see [**8**, Proposition 3.6, pp. 77–78] or [**7**, pp. 323–324]. We have motivated (19) here because it is useful in the proof of the main Theorem 3. Notice that the construction in (19) is valid for *any* vector field η, although $\eta = f + g$ is the immediate interest.

Proposition 1. *The tangent vector $V'(0)$ of the curve* (19) (*where $\eta = f+g$*) *is*

$$V'(0) = [f,\eta](x_0) = [f,g](x_0).$$

If f and g are analytic, then $V^{(k)}(0) = ad_f^k g(x_0)$ for all $k \geq 1$.

Proof: Use the formula for the derivative of a product, together with the variational equations, to compute

$$\begin{aligned} V'(t) &= \frac{\partial}{\partial t}(\phi_{-t}^f)_*\eta(\phi_t^f(x_0)) + (\phi_{-t}^f)_*\frac{\partial\eta}{\partial x}\frac{\partial}{\partial t}(\phi_t^f(x_0)) \\ &= -(\phi_{-t}^f)_*\frac{\partial f}{\partial x}\eta(\phi_t^f(x_0)) + (\phi_{-t}^f)_*\frac{\partial\eta}{\partial x}f(\phi_t^f(x_0)) \\ &= (\phi_{-t}^f)_*[f,\eta](\phi_t^f(x_0)). \end{aligned} \tag{20}$$

Set $t = 0$ to get $V'(0) = [f,\eta](x_0)$, taking into account the initial conditions $\phi_0^f(x) = x$ and $(\phi_0^f)_* = I$. Since $[f, f+g] = [f,g]$, the first statement is proved. Given (20), the second statement follows by induction. ∎

Because of Proposition 1, the Lie bracket $[f, \eta]$ is also called the *Lie derivative of η along f*.

The algebraic property that shows the connection between the Lie bracket operation on vector fields and the Lie derivative operation on functions is the *Jacobi identity*; it says that if v, w are vector fields and λ is a smooth function, then $(L_{[v,w]}\lambda)(x) = (L_v L_w \lambda - L_w L_v \lambda)(x)$. This identity is proved as follows. For any smooth λ, and any x,

$$\begin{aligned} L_v L_w \lambda(x) - L_w L_v \lambda(x) &= L_v(d\lambda(x) \cdot w(x)) - L_w(d\lambda(x) \cdot v(x)) \\ &= \left(d^2\lambda(x) \cdot w(x) + d\lambda(x) \cdot \frac{\partial w}{\partial x}(x)\right) \cdot v(x) \\ &\quad - \left(d^2\lambda(x) \cdot v(x) + d\lambda(x) \cdot \frac{\partial v}{\partial x}(x)\right) \cdot w(x), \quad (21) \end{aligned}$$

where $d^2\lambda(x)$ is the matrix of second partial derivatives of λ at x. Since $d^2\lambda$ is symmetric, we get

$$L_v L_w \lambda - L_w L_v \lambda = d\lambda \cdot \left(\frac{\partial w}{\partial x} v - \frac{\partial v}{\partial x} w\right) = d\lambda \cdot [v, w] = L_{ad_v w} \lambda.$$

The Jacobi identity itself helps to identify $\ker d\lambda(x)$ for $x \in U$, for a function λ having relative degree n. Given g as the replacement for b in (5), and assuming that $d\lambda(x) \cdot g(x) = 0$, one can show that $ad_f^k g(x) \in \ker d\lambda(x)$ for $k = 1, \ldots n-2$ by induction, using the Jacobi identity. We now give the details of the nonlinear version of the calculation in (5).

Proposition 2. *If λ has relative degree n with respect to* (6a) *in the open set U, then for all $x \in U$,*

(1) *the covectors $d\lambda(x)$, $dL_f\lambda(x), \ldots, dL_f^{n-1}\lambda(x)$ are linearly independent*;
(2) *the vectors $g(x)$, $ad_f g(x), \ldots, ad_f^{n-1} g(x)$ are linearly independent.*

Proof: Consider the matrix product that generalizes (5):

$$\begin{bmatrix} d\lambda(x) \\ dL_f\lambda(x) \\ \vdots \\ dL_f^{n-1}\lambda(x) \end{bmatrix} \begin{bmatrix} g(x) & ad_f g(x) \ldots ad_f^{n-1} g(x) \end{bmatrix}$$

$$= \begin{bmatrix} L_g\lambda(x) & L_{ad_f g}\lambda(x) & \cdots & \cdots & L_{ad_f^{n-1} g}\lambda(x) \\ L_g L_f \lambda(x) & & & L_{ad_f^{n-2} g} L_f \lambda(x) & \star \\ \vdots & & & & \vdots \\ L_g L_f^{n-1}\lambda(x) & \star & \cdots & \cdots & \star \end{bmatrix}. \quad (22)$$

We now use the relative degree n assumption and the Jacobi identity to show that the matrix on the right in (22) is lower right triangular with nonzero entries on the skew-diagonal; Proposition 2 then follows.

By relative degree n, the first column has the required form. Proceed by induction on the columns, using the Jacobi identity. Now the k, l entry in the matrix is $(dL_f^k h) \cdot (ad_f^l g)$, for $0 \le k, l \le n-1$. The diagonal entries in question

are those for which $k + l = n - 1$. Assume the desired property for column l: thus, assume that $L_{ad_f^l g} L_f^k h = 0$ for $k + l \le n - 2$, and $L_{ad_f^l g} L_f^{n-1-l} h \ne 0$. For column $l + 1$ we need $L_{ad_f^{l+1} g} L_f^k h = 0$ for $k \le n - 3 - l$, and $L_{ad_f^{l+1} g} L_f^{n-2-l} h \ne 0$. Using the Jacobi identity, column $l + 1$ is, for $0 \le k \le n - 1$:

$$(k, l) \text{ } entry = \left(dL_f^k h\right) \cdot \left(ad_f^{l+1} g\right) = L_f L_{ad_f^l g} L_f^k h - L_{ad_f^l g} L_f^{k+1} h. \tag{23}$$

Apply the hypothesis to (23) for $k = 0, \ldots, n - 3 - l$ to get zero. For $k = n - 2 - l$, only the last term in (23) contributes to the skew-diagonal entry, which is

$$-L_{ad_f^l g} L_f^{n-1-l} h \ne 0.$$

Notice that this is the *negative* of the skew-diagonal entry in column l. Since the first column has last entry $L_g L_f^{n-1} h(x)$, the diagonal entry in column l must therefore be $(-1)^l L_g L_f^{n-1} h(x) \ne 0$ for $l = 0, \ldots, n - 1$. Thus, our matrix is lower right triangular with nonzero skew-diagonal entries for $x \in U$. ■

Notice that Proposition 2 completes the proof of Theorem 2, because it shows that if λ has relative degree n then the vector function

$$\left[\lambda(x), L_f \lambda(x), \ldots, L_f^{n-1} \lambda(x)\right]^T$$

has a nonsingular Jacobian for $x \in U$.

As in (5), a geometric interpretation of Proposition 2 is that the null space of the differential $d\lambda(x)$ is the $(n - 1)$-dimensional space

$$\mathscr{D}(x) = span\left\{g(x), ad_f g(x), \ldots, ad_f^{n-2} g(x)\right\}, \quad x \in U. \tag{24}$$

We view each subspace $\mathscr{D}(x)$ as a subspace of the tangent space of R^n at x, $T_x R^n \approx R^n$. The collection of the subspaces (24) for $x \in U$ is called a *distribution* on U. In the linear case, a constant distribution such as $span\{b, Ab, \ldots, A^{n-2} b\}$ is automatically the null space of a nonzero linear functional. Proposition 2 shows that the nonsingularity of both factors on the left in (22) is *necessary* for the solution of the ISLP. Under the assumption of Proposition 2, there is an additional necessary condition on the distribution (24), which is not revealed in the calculations of Proposition 2. The additional condition on $\mathscr{D}(x)$ is the geometric condition of *involutivity.* Involutivity is an integrability condition that guarantees that the distribution (24) is the space annihilated by the differential of a function having relative degree n. The next section examines this concept, and develops the geometric conditions for the solvability of the ISLP.

4. GEOMETRIC CRITERIA FOR ISLP SOLVABILITY. It's convenient to place some formal definitions here.

Definition 3. A distribution $\mathscr{D}$ on U is a smooth assignment (via functions like g, $ad_f g$, etc. in (24)) of a subspace of the tangent space $T_x U \approx R^n$, for $x \in U$. A distribution $\mathscr{D}(x)$ is *involutive* in U if, for vector fields v_1 and v_2, and all $x \in U$,

$$v_1(x), v_2(x) \in \mathscr{D}(x) \Rightarrow [v_1, v_2](x) \in \mathscr{D}(x).$$

A distribution $\mathscr{D}$ is *nonsingular* in U if dim $\mathscr{D}(x)$ is constant in U. A nonsingular distribution $\mathscr{D}$ with dim $\mathscr{D}(x) = k$ is *integrable* in U if there are $n - k$ functions λ_j such that $span\{d\lambda_j(x) : 1 \le j \le n - k\} = \mathscr{D}^\perp(x)$, or equivalently, $\bigcap_{i=1}^{n-k} \ker d\lambda_j(x) = \mathscr{D}(x)$.

To illustrate the involutivity concept, we return to Example 1.

Example 7. The distribution (24) for Example 1 is $\mathscr{D}(x) = span\{g(x), ad_f g(x)\}$. Appropriate bracket calculations for Example 1 give

$$ad_f g(x) = -\begin{bmatrix} 0 & -e^{-x_2} & 0 \\ 1 & 0 & 0 \\ x_1 & 0 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ x_1 \end{bmatrix},$$

and then

$$[g, ad_f g](x) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

If we form the matrix

$$\begin{bmatrix} g(x) & ad_f g(x) & [g, ad_f g](x) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & x_1 & 1 \end{bmatrix},$$

we see that its rank is everywhere equal to 3, so that $[g, ad_f g](x)$ does not lie in $\mathscr{D}(x)$ for any x. It follows that the distribution $\mathscr{D} = span\{g, ad_f g\}$ is not involutive in any open set. Compare this situation with Example 5.

Frobenius' Theorem states that for a nonsingular distribution $\mathscr{D}$ (of *any* dimension), integrability is *equivalent* to involutivity [**4**, p. 23]. The following discussion of the ideas of Frobenius' Theorem in the case of interest for the Input-State Linearization Problem uses virtually all the tools discussed so far.

It is easy to show that involutivity is *necessary* for $\mathscr{D}(x)$ in (24) to be the tangent space at x of the level set within U of a smooth function λ having relative degree n. If λ has relative degree n at x_0, Proposition 2 shows that $[g(x)\, ad_f g(x) \dots ad_f^{n-1} g(x)]$ has rank n for all $x \in U$, so the distribution $\mathscr{D}$ is nonsingular with constant dimension $n - 1$ in U. Moreover, we have $d\lambda(x)[g(x)\, ad_f g(x) \dots ad_f^{n-2} g(x)] = 0$. Thus, the pointwise orthogonal complement of $\mathscr{D}$ in U is $\mathscr{D}^{\perp} = span\{d\lambda\}$. This says that the distribution $\mathscr{D}$ is integrable on U. We now show involutivity directly. If $0 \le i, j \le n - 2$, then for $x \in U$,

$$d\lambda(x) \cdot \left[ad_f^i g, ad_f^j g\right](x) = L_{ad_f^i g} L_{ad_f^j g} \lambda(x) - L_{ad_f^j g} L_{ad_f^i g} \lambda(x) = 0$$

since λ has relative degree n. Therefore $[ad_f^i g, ad_f^j g](x) \in \mathscr{D}(x)$ for all x in U. This is sufficient to show that $\mathscr{D}$ is involutive, since any two vector fields in $\mathscr{D}$ are linear combinations of the ones we just dealt with, and one can show that for fields ξ, η in $\mathscr{D}$ and functions a, b, we have

$$[a\xi, b\eta](x) = a(x)b(x)[\xi, \eta](x) + \big(L_\xi b(x)\big)a(x)\eta(x) - \big(L_\eta a(x)\big)b(x)\xi(x). \tag{25}$$

To establish (25), use the Jacobi identity plus the characterization that vector fields v, w are equal if and only if $L_v \lambda = L_w \lambda$ for all smooth functions λ.

Theorem 3 provides the goal of computable geometric criteria on f and g for the solution of the Input-State Linearization Problem. Involutivity is the necessary *and* sufficient integrability condition for the system of partial differential equations (16a).

Theorem 3. *The system $x' = f(x) + g(x)u$ is input-state linearizable in a neighborhood U of x_0 if and only if*

(I) $[g(x_0)\; ad_f g(x_0) \dots ad_f^{n-1} g(x_0)]$ *has rank n, and*
(II) *the distribution $\mathscr{D} = span\{g, ad_f g, \dots, ad_f^{n-2} g\}$ is involutive in U.*

Proof: We show that (I) and (II) are equivalent to the existence of a function λ having relative degree n at x_0. *Necessity* of (I) is covered by Proposition 2, and necessity of (II) was discussed before the theorem statement.

Sufficiency. If (I) and (II) hold, then the distribution $\mathscr{D}$ is nonsingular on U with dimension $n-1$. Clearly, there is a smooth covector field in U defined by a smooth row vector $w(x)$ such that $w(x) = span\,\mathscr{D}^{\perp}(x)$ and also $w(x) \cdot ad_f^{n-1} g(x) \neq 0$ for $x \in U$. Thus, for all $x \in U$, we have

$$span\left\{g(x), \dots, ad_f^{n-2} g(x), w(x)\right\} = R^n.$$

The function λ can be constructed from the flows of these vector fields. To simplify notation for this argument, define $v_1 = g$, $v_2 = ad_f g, \dots, v_{n-1} = ad_f^{n-2} g$, $v_n = w$.

Let U_ϵ be a ball of radius ϵ about the zero vector in R^n. There is an $\epsilon > 0$ such that the map $\psi : U_\epsilon \to \psi(U_\epsilon) \subset U$ defined by

$$\psi(z) = \psi(z_1, \dots, z_n) = \phi_{z_1}^{v_1} \circ \phi_{z_2}^{v_2} \circ \cdots \circ \phi_{z_n}^{v_n}(x_0)$$

is a diffeomorphism onto its image, that is, ψ is smooth, one-to-one, and has a smooth inverse map defined on $\psi(U_\epsilon)$. This is because repeated application of the chain rule shows that at $z = 0$ we have

$$\frac{\partial \psi}{\partial z_i}(0) = v_i(x_0), \quad i = 1, \dots, n, \tag{26}$$

and by hypothesis, the vectors $v_i(x_0)$ are independent. Thus, the Inverse Function Theorem ensures that there is an $\epsilon > 0$ so that ψ is a local diffeomorphism onto its image. The z coordinates are time coordinates that "straighten out" the flows for the vector fields v_i. Write the inverse of ψ in the form

$$\psi^{-1}(x) = \begin{bmatrix} \lambda_1(x) \\ \vdots \\ \lambda_n(x) \end{bmatrix}.$$

Now consider (26) together with the identity

$$\left(\frac{\partial \psi^{-1}}{\partial x}\right)_{z = \psi^{-1}(x)} \left(\frac{\partial \psi}{\partial z}\right)_{x = \psi(z)} = I_{n \times n}. \tag{27}$$

The strategy is to show that $d\lambda_n(x)$ spans $\mathscr{D}^{\perp}(x)$, by showing that the first $n-1$ columns of $(\partial \psi / \partial z)_{x = \psi(z)}$ form a basis of $\mathscr{D}(x)$ at any $x \in U$, for then (27) implies that $span\{d\lambda_n(x)\} = \mathscr{D}^{\perp}(x)$ for $x \in U$.

Using the chain rule, we find that the i-th column of $\partial \psi / \partial z$ is

$$\begin{aligned}
\frac{\partial \psi}{\partial z_i} &= \left(\phi_{z_1}^{v_1}\right)_* \circ \cdots \circ \left(\phi_{z_{i-1}}^{v_{i-1}}\right)_* \frac{\partial}{\partial z_i}\left(\phi_{z_i}^{v_i} \circ \cdots \circ \phi_{z_n}^{v_n}(x_0)\right) \\
&= \left(\phi_{z_1}^{v_1}\right)_* \circ \cdots \circ \left(\phi_{z_{i-1}}^{v_{i-1}}\right)_* v_i\left(\phi_{z_i}^{v_i} \circ \cdots \circ \phi_{z_n}^{v_n}(x_0)\right) \\
&= \left(\phi_{z_1}^{v_1}\right)_* \circ \cdots \circ \left(\phi_{z_{i-1}}^{v_{i-1}}\right)_* v_i\left(\phi_{-z_{i-1}}^{v_{i-1}} \circ \cdots \circ \phi_{-z_1}^{v_1}(\psi(z))\right),
\end{aligned}$$

where we use the fact that $(\phi_z^{v_i})^{-1} = \phi_{-z}^{v_i}$ for the local flows of the v_i. If we show that, whenever ξ, η are vector fields that are pointwise in $\mathscr{D}$ we also have

$$(\phi_t^\eta)_* \xi(\phi_{-t}^\eta(x)) \in \mathscr{D}(x),$$

then it follows that the column vector $(\partial\psi/\partial z_i)_{z=\psi^{-1}(x)}$ is also in $\mathscr{D}(x)$. Since any vector field ξ in $\mathscr{D}$ can be written as $\Sigma c_i v_i$ for some functions c_i in $\mathscr{D}$, we need only consider the case when $\xi = v_i$ for some $i = 1, \ldots, n-1$.

Thus, for a vector field $\eta \in \mathscr{D}$, let x be *a fixed point* in $\mathscr{D}$, and set

$$V_i(t) = (\phi_{-t}^\eta)_* v_i(\phi_t^\eta(x)); \quad i = 1, \ldots, n-1. \tag{28}$$

The vector functions V_i are defined for some interval of t about 0. It follows from Proposition 1 that

$$\frac{d}{dt}V_i(t) = (\phi_{-t}^\eta)_*[\eta, v_i](\phi_t^\eta(x)).$$

Since $\mathscr{D}$ is involutive and $\eta, v_i \in \mathscr{D}$, there exist functions α_{ij} defined around x such that

$$[\eta, v_i] = \sum_{j=1}^{n-1} \alpha_{ij} v_j,$$

so that

$$\frac{d}{dt}V_i(t) = (\phi_{-t}^\eta)_* \left(\sum_{j=1}^{n-1} \alpha_{ij} v_j \right)(\phi_t^\eta(x)) = \sum_{j=1}^{n-1} \alpha_{ij}(\phi_t^\eta(x)) V_j(t)\ .$$

Thus, $V = [V_1 \ldots V_{n-1}]$ is a matrix solution of a linear system of differential equations having the form $V' = VA^T$, where $A \equiv [\alpha_{ij}]$ and $1 \le i, j \le n-1$. Therefore we can write

$$[V_1(t) \ldots V_{n-1}(t)] = [V_1(0) \ldots V_{n-1}(0)]X(t), \tag{29}$$

where $X(t)$ is an $(n-1) \times (n-1)$ fundamental matrix of solutions. Multiply (29) from the left by $(\phi_t^\eta)_*$ and use (28) to get

$$[v_1(\phi_t^\eta(x)) \ldots v_{n-1}(\phi_t^\eta(x))] = [(\phi_t^\eta)_* v_1(x) \ldots (\phi_t^\eta)_* v_{n-1}(x)]X(t).$$

Now, for small t we may replace x by $\phi_{-t}^\eta(x)$ on the same orbit to get

$$[v_1(x) \ldots v_{n-1}(x)] = [(\phi_t^\eta)_* v_1(\phi_{-t}^\eta(x)) \ldots (\phi_t^\eta)_* v_{n-1}(\phi_{-t}^\eta(x))]X(t).$$

Since $X(t)$ is nonsingular, we get for $i = 1, \ldots, n-1$,

$$\begin{aligned}(\phi_t^\eta)_* v_i(\phi_{-t}^\eta(x)) &\in span\{v_1(x), \ldots, v_{n-1}(x)\} \\ &= span\{g(x), \ldots, ad_f^{n-2}g(x)\} = \mathscr{D}(x).\end{aligned}$$

The proof that the columns $\partial\psi/\partial z_i$ are in $\mathscr{D}$ for $i = 1, \ldots, n-1$ is now complete.

It remains to show that λ_n satisfies the nontriviality condition required for relative degree n. But from the identity (27) and the conditions (26) we have

$$d\lambda_n(x_0) \cdot \frac{\partial\psi}{\partial z_n}(0) = d\lambda_n(x_0) \cdot v_n(x_0) = d\lambda_n(x_0) \cdot w(x_0) = 1.$$

Thus, $d\lambda_n(x_0)$ is parallel to $w(x_0) \in \mathscr{D}^\perp(x_0)$, so $d\lambda_n(x_0) \cdot ad_f^{n-1}g(x_0) \neq 0$. Now apply the Jacobi identity to

$$L_{ad_f^{n-1}g}\lambda_n(x_0) = L_{[f, ad_f^{n-2}g]}\lambda_n(x_0),$$

to conclude that $-L_{ad_f^{n-2}g} L_f \lambda_n(x_0) = (-1)^{n-1} L_g L_f^{n-1} \lambda(x_0) \neq 0$. Therefore λ_n has relative degree n in some open set in U. ■

The proof of sufficiency in Theorem 3 follows the proof in [**4**, pp. 24–28] that involutivity implies integrability. We really needed a particular *codimension one* case of Frobenius' Theorem, where dim $\mathscr{D}^{\perp} = 1$ and $\mathscr{D}$ is the special distribution in (24); however, the argument for the general case is much the same.

Returning to Example 3, conditions (I) and (II) of Theorem 3 are easily checked: (I) holds near $x_0 = 0$, and (II) holds trivially since $\mathscr{D}(x) = span\{g(x)\} = span\{[0\ 1]^T\}$ is one-dimensional. In fact, in dimension $n = 2$, conditions (I) and (II) reduce to the single condition that

$$rank\left[g(x_0)[f,g](x_0)\right] = 2,$$

because then $g(x) \neq 0$ in some open set about x_0.

Example 8. The result of Example 7 showed that condition (II) of Theorem 3 does not hold for the system of Example 1. Let us check that condition (I) does hold: we just need to compute

$$ad_f^2 g(x) = \left[f, ad_f g\right](x) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}\begin{bmatrix} e^{-x_2} \\ x_1 \\ \frac{1}{2}x_1^2 \end{bmatrix} - \begin{bmatrix} 0 & -e^{-x_2} & 0 \\ 1 & 0 & 0 \\ x_1 & 0 & 0 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \\ x_1 \end{bmatrix};$$

then, using the calculations from Example 7, we have

$$\left[g(x) \quad ad_f g(x) \quad ad_f^2 g(x)\right] = \begin{bmatrix} 1 & 0 & e^{-x_2} \\ 0 & 1 & 0 \\ 0 & x_1 & e^{-x_2} \end{bmatrix},$$

so the rank condition (I) is satisfied at every point.

We can now discuss condition (I) of Theorem 3 as a type of controllability condition known as a *local reachability* condition.

Suppose $f(x_0) = 0$, so that x_0 is an equilibrium for $x' = f(x)$. The rank condition (I) is exactly the condition that the local linearization defined by

$$x' = Ax + bu \equiv \frac{\partial f}{\partial x}(x_0)x + g(x_0)u \tag{30}$$

is completely controllable, that is, $rank[bAb \ldots A^{n-1}b] = n$. To see this, notice that if we write

$$f(x) = Ax + f_2(x), \quad g(x) = b + g_1(x),$$

where $(\partial f_2/\partial x)(x_0) = 0$ and $g_1(x_0) = 0$, then an induction proof shows that for each k,

$$ad_f^k g(x) = (-1)^k A^k b + p_k(x), \quad p_k(x_0) = 0. \tag{31}$$

Indeed, (31) holds when $k = 0$ because $g(x) = b + g_1(x)$ with $g_1(x_0) = 0$. The induction step follows by an appropriate bracket calculation using the expansions for f and g. Thus, (I) is exactly the complete controllability condition for (30), provided $f(x_0) = 0$.

Let us say that a system $x' = f(x) + g(x)u$ is *locally reachable* at x_0 if there is an open set U about x_0 such that every point $x \in U$ can be reached from x_0 in

finite time by means of a control $u(t)$. Under the conditions of Theorem 3, the nonlinear system is locally reachable at x_0 because the system is locally equivalent to the completely controllable linear system (12), via the coordinate mapping $z = T(x)$ for $x \in U$ (and $z \in T(U)$), and the feedback transformation $u = \alpha(x) + \beta(x)v = \alpha(T^{-1}z) + \beta(T^{-1}z)v$. For if a control $v(t)$ transfers $z_0 = 0$ to the point z_f in time t_f while the trajectory $z(t)$ remains in $T(U)$ (and this can be done for system (12)), then the corresponding $u(t)$ keeps $x(t) = T^{-1}(z(t))$ within U. Therefore, transfers from x_0 to any x_f in U can be accomplished in finite time. If only the rank condition (I) holds, however, then provided $f(x_0) = 0$, local reachability of the nonlinear system at x_0 can still be proved using the controllability rank condition for the local linearization (30) together with the help of the Inverse Function Theorem. For one result of this type, which implies local reachability at x_0, see [**8**, Proposition 3.3, pp. 74–75].

The method of input-state linearization has been successful in addressing specific control problems in the areas of aircraft flight control and robotics [**10**, p. 207]. Here is a final example concerning the equations for a single-link robotic manipulator.

Example 9. [**6**, p. 528] [**10**, p. 242] The dynamical equations for a single-link, flexible-joint mechanism with negligible damping are

$$\begin{aligned} Iq_1'' + MGL\sin q_1 + k(q_1 - q_2) &= 0 \\ Jq_2'' - k(q_1 - q_2) &= u, \end{aligned}$$

where q_1 and q_2 are angular positions, I and J are moments of inertia, k is a spring constant, M is a mass, G is the gravitational constant, L is a distance, and u is a motor torque input. By writing $x = [q_1\ q_1'\ q_2\ q_2']^T$, the four-dimensional state equations can be written as

$$x' = f(x) + g(x)u \equiv \begin{bmatrix} x_2 \\ -a\sin x_1 - b(x_1 - x_3) \\ x_4 \\ c(x_1 - x_3) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ d \end{bmatrix} u, \tag{32}$$

using these positive constants: $a = (MGL)/I$, $b = k/I$, $c = k/J$, and $d = 1/J$. The unforced system has an equilibrium at $x = 0$. To determine if this system is input-state linearizable near the origin, we check conditions (I), (II) of Theorem 3. First, compute

$$\frac{\partial f}{\partial x} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -a\cos x_1 - b & 0 & b & 0 \\ 0 & 0 & 0 & 1 \\ c & 0 & -c & 0 \end{bmatrix}.$$

In this case, the spanning vector fields for $\mathscr{D}$ are constant, and appropriate bracket calculations lead to the matrix needed in condition (I):

$$\begin{bmatrix} g & ad_f g & ad_f^2 g & ad_f^3 g \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & -bd \\ 0 & 0 & bd & 0 \\ 0 & -d & 0 & cd \\ d & 0 & -cd & 0 \end{bmatrix}.$$

Since this matrix has rank 4, condition (I) holds everywhere. It is immediate that the involutivity condition (II) holds everywhere, since $\mathscr{D}(x) = span\{g, ad_f g, ad_f^2 g\}$

is spanned by constant vector fields. By Theorem 3, system (32) is input-state linearizable, and the construction of the function T_1 with relative degree 4 can now be attempted. In fact, since we know that the null space of $dT_1(x)$ must be the flat distribution (linear subspace) defined by $span\{g, ad_f g, ad_f^2 g\}$, we take T_1 to be a linear function of x_1 alone: $T_1(x) = x_1$. The complete coordinate transformation $T(x)$ is then obtained from (16a), yielding

$$z = T(x) = \begin{bmatrix} T_1(x) \\ T_2(x) \\ T_3(x) \\ T_4(x) \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ -a \sin x_1 - b(x_1 - x_3) \\ -ax_2 \cos x_1 - b(x_2 - x_4) \end{bmatrix}.$$

Alternatively, one can construct T_1 directly from (16) for this example: the details appear in [**6**, pp. 528–529]. By defining the feedback transformation $u = \alpha(x) + \beta(x)v$ according to (17), the equations for z are given by (12); equivalently, $z_1^{(4)} = v$. A control v can now be designed that makes the link position $x_1 = z_1$ track a prespecified trajectory.

Clearly, systems that are fully input-state linearizable are very special. Nevertheless, as shown by the analysis in [**4**, pp. 162 – 172], if a system is not fully input-state linearizable as in Theorem 3, but an output function with relative degree $r < n$ is known, then the system can still be transformed to a partially linear normal form, which is quite useful in many control problems. As we noted in Example 3, the case when $r < n$ requires the examination of unobservable dynamics (usually called *zero dynamics*) [**4**, pp. 163–164]. However, the circle of ideas discussed here can be applied to a wide range of problems beyond the special conditions of Theorem 3.

9. FURTHER READING. The equivalence problem discussed here was first considered in [**1**]. For the origin of the use of Lie brackets in the study of reachability problems see [**3**] and the references therein. See also [**5**, pp. 1–2] for some historical insight on the introduction into control theory of differential-geometric ideas centered around the Lie bracket. Reference [**4**] presents important control methods using extensions of the basic ideas discussed in this article, and includes a development of the required differential-geometric concepts. References [**4**] and [**10**] discuss the important issue of stability of unobservable dynamics (called *zero dynamics*) that arose in Example 3, which comes from [**6**]. Information on controllability, observability, and numerous other issues appears in [**8**] and [**11**]. Reference [**8**, p. 59] contains a statement of the classical Frobenius' Theorem on integrability of a system of linear, first-order partial differential equations. See [**6**] and [**7**] for some engineering emphasis combined with excellent theoretical exposition. Much of the differential-geometric nonlinear control theory generalizes the geometric approach to linear control theory presented in [**13**]. Many references to the primary mathematical control literature may be found in the references listed.

REFERENCES

1. R. W. Brockett, Feedback invariants for nonlinear systems, in *Proceedings International Federation on Automatic Control Congress*, Helsinki, 1978, pp. 1115–1120.
2. J. K. Hale, *Ordinary Differential Equations*, Second Edition, Krieger, Malabar, 1980.
3. R. Hermann, "On the accessibility problem in control theory," in *International Symposium on Nonlinear Differential Equations and Nonlinear Mechanics* (eds. J. P. Lasalle, S. Lefschetz), Academic Press, New York, 1963, pp. 325–332.

4. A. Isidori, *Nonlinear Control Systems*, Third Edition, Springer, New York, 1995.
5. V. Jurdjevic, *Geometric Control Theory*, Cambridge Studies in Advanced Mathematics 51, Cambridge University Press, Cambridge, 1997.
6. H. K. Khalil, *Nonlinear Systems*, Second Edition, Prentice-Hall, Inc., Englewood Cliffs, 1996.
7. R. M. Murray, Z. Li, and S. S. Sastry, *A Mathematical Introduction to Robotic Manipulation*, CRC Press, Boca Raton, 1994.
8. H. Nijmeijer and A. J. van der Schaft, *Nonlinear Dynamical Control Systems*, Springer-Verlag, New York, 1990.
9. W. Rudin, *Principles of Mathematical Analysis*, Second Edition, McGraw-Hill, New York, 1964.
10. J. E. Slotine and W. Li, *Applied Nonlinear Control*, Prentice-Hall, Inc., Englewood Cliffs, 1991.
11. E. D. Sontag, *Mathematical Control Theory: Deterministic Finite Dimensional Systems*, Second Edition, Springer-Verlag, New York, 1998.
12. W. J. Terrell, Some Fundamental Control Theory I: Controllability, Observability, and Duality, *Amer. Math. Monthly*, to appear.
13. W. M. Wonham, *Linear Multivariable Control*, Third Edition, Springer-Verlag, New York, 1985.

WILLIAM J. TERRELL received his Ph.D. in Applied Mathematics from North Carolina State University in 1990. He is an Associate Professor in the Department of Mathematical Sciences at Virginia Commonwealth University, Richmond, Virginia. His research interests are in the area of differential equations with special focus on systems and control theory.
Department of Mathematical Sciences, Virginia Commonwealth University, Richmond, VA 23284-2014
wterrell@atlas.vcu.edu

MATHEMATICS

C. O. Oakley, Haverford College

Mathematics is one component of any plan for liberal education. Mother of all the sciences, it is a builder of the imagination, a weaver of patterns of sheer thought, an intuitive dreamer, a poet. The study of mathematics cannot be replaced by any other activity that will train and develop man's purely logical faculties to the same level of rationality. Through countless dimensions, riding high the winds of intellectual adventure and filled with the zest of discovery, the mathematician tracks the heavens for harmony and eternal verity. There is not wholly unexpected surprise, but surprise nevertheless, that mathematics has direct application to the physical world about us. For mathematics, in a wilderness of tragedy and change, is a creature of the mind, born to the cry of humanity in search of an invariant reality, immutable in substance, unalterable with time. Mathematics is an infinity of flexibles forcing pure thought into a cosmos. It is an arc of austerity cutting realms of reason with geodesic grandeur. Mathematics is crystallized clarity, precision personified, beauty distilled and rigorously sublimated. The life of the spirit is a life of thought; the ideal of thought is truth; everlasting truth is the goal of mathematics.

MONTHLY 56 (1949) 19

The Dirichlet Problem for Ellipsoids

John A. Baker

The purpose of this paper is to present two elementary (and perhaps somewhat novel) solutions of the Dirichlet problem for ellipsoids in $\mathbb{R}^n$. One of these is based on an elegant result of Ernst Fischer—of Riesz–Fischer fame.

By the *Dirichlet problem* (for the Laplacian) we mean the following: Given a bounded region (nonempty, open, connected set) Ω in $\mathbb{R}^n, n \geq 2$, and given a continuous function $f: \partial\Omega \to \mathbb{R}$ (called the *boundary data*), find a continuous function $u: \overline{\Omega} \to \mathbb{R}$ such that u is C^2 on Ω,

$$\sum_{k=1}^{n} \frac{\partial^2}{\partial x_k^2} u(x) = 0 \quad \text{for} \quad x \in \Omega \quad \text{and}$$

$$u(x) = f(x) \quad \text{for} \quad x \in \partial\Omega\text{—the boundary of } \Omega.$$

This is surely one of the most influential problems for the development of mathematics in the last two centuries; see, for example, [**4**], [**6**], and [**8**]. The case in which Ω is a disk in $\mathbb{R}^2$ is standard fare for writers of texts on the theory of analytic functions of one complex variable. In this paper we are concerned with the case in which Ω is an ellipsoid in $\mathbb{R}^n$, for arbitrary $n \geq 2$, and especially when f is (the restriction to $\partial\Omega$ of) a polynomial function.

1. BACKGROUND AND NOTATION. Let $2 \leq n \in \mathbb{Z}$. For $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ in $\mathbb{R}^n$, let $x \cdot y = \sum_{k=1}^{n} x_k y_k$ and $|x| = (x \cdot x)^{1/2}$. For $1 \leq k \leq n$ we write ∂_k instead of $\partial / \partial x_k$ and define the *Laplacian* $\Delta = \sum_{k=1}^{n} \partial_k^2$. If Ω is a region in $\mathbb{R}^n$ and $v: \Omega \to \mathbb{R}$ then we say that v is *harmonic* on Ω provided v is C^2 (twice continuously differentiable) on Ω and v satisfies the *Laplace equation*

$$\Delta v(x) = 0 \quad \text{for all } x \in \Omega.$$

For completeness we include a well known proof [**9**, p. 103] of the weak form of

The Maximum Principle. *If Ω is a bounded region in $\mathbb{R}^n$, $v: \overline{\Omega} \to \mathbb{R}$, v is continuous on $\overline{\Omega}$, v is C^2 on Ω, and $\Delta v(x) \geq 0$ for all $x \in \Omega$, then v attains its maximum on $\partial\Omega$.*

Proof: For $0 < \epsilon \in \mathbb{R}$ let $v_\epsilon(x) = v(x) + \epsilon |x|^2$ for $x \in \overline{\Omega}$. Then v_ϵ is continuous on $\overline{\Omega}$, C^2 on Ω and, for all $x \in \Omega$,

$$\sum_{k=1}^{n} \partial_k^2 v_\epsilon(x) = \sum_{k=1}^{n} \partial_k^2 v(x) + 2n\epsilon \geq 2n\epsilon > 0.$$

Hence, for each $x \in \Omega$ there is a $k (1 \leq k \leq n)$ such that $\partial_k^2 v_\epsilon(x) > 0$; single variable calculus ensures that v_ϵ does *not* have a relative maximum at x. It follows that, for each $\epsilon > 0, v_\epsilon$ attains its maximum on $\partial\Omega$. That is, for each $\epsilon > 0$ there exists $x_\epsilon \in \partial\Omega$ such that $v_\epsilon(x) \leq v_\epsilon(x_\epsilon)$ for all $x \in \overline{\Omega}$. Hence, for $\epsilon > 0$ and $x \in \overline{\Omega}$ $v(x) \leq v_\epsilon(x) \leq v(x_\epsilon) + \epsilon |x_\epsilon|^2$, so that $v(x) \leq \max\{v(y): y \in \partial\Omega\} + \epsilon R^2$, where $R = \max\{|x| : x \in \overline{\Omega}\}$. Since this is so for every $\epsilon > 0$, $v(x) \leq \max\{v(y): y \in \partial\Omega\}$ for all $x \in \overline{\Omega}$. ∎

Corollary 1. *If Ω is a bounded region in $\mathbb{R}^n$, $u : \overline{\Omega} \to \mathbb{R}$, and u is continuous on $\overline{\Omega}$ and harmonic on Ω, then u attains its maximum and its minimum on $\partial\Omega$.*

For proof it suffices to note that since u is harmonic on Ω, so is $-u$.

Corollary 2. *If Ω is a bounded region in $\mathbb{R}^n$ and $f : \partial\Omega \to \mathbb{R}$ is continuous, then there is at most one solution to the Dirichlet problem satisfying $u(x) = f(x)$ for all $x \in \partial\Omega$.*

Proof: Suppose that u_1 and u_2 were such solutions and let $v(x) = u_1(x) - u_2(x)$ for $x \in \overline{\Omega}$. Then v is continuous on $\overline{\Omega}$, harmonic on Ω, and vanishes on $\partial\Omega$. Hence, by Corollary 1, for every $x \in \overline{\Omega}$ we have

$$0 = \min\{v(y) : y \in \partial\Omega\} \le v(x) \le \max\{v(y) : y \in \partial\Omega\} = 0,$$

i.e., $0 = v(x) = u_1(x) - u_2(x)$ for all $x \in \overline{\Omega}$. ∎

Let's fix $n \ge 2$ in $\mathbb{Z}$, let $r_1, \dots, r_n > 0$, let

$$b(x) = 1 - \sum_{k=1}^{n} x_k^2/r_k^2 \qquad \text{for } x = (x_1, \dots, x_n) \in \mathbb{R}^n$$

and let $\mathscr{E}$ denote the ellipsoid $\{x \in \mathbb{R}^n : b(x) > 0\}$ so that $\partial\mathscr{E} = \{x \in \mathbb{R}^n : b(x) = 0\}$. We aim to solve the Dirichlet problem for $\mathscr{E}$ by first showing that it can be handled in the case of polynomial boundary data with the aid of an elementary consequence of some work of E. Fischer.

Let $\mathscr{P}$ denote the real algebra of all polynomial functions from $\mathbb{R}^n$ into $\mathbb{R}$. For $0 \le m \in \mathbb{Z}$, $\mathscr{P}_m$ denotes the *finite dimensional* linear subspace of $\mathscr{P}$ consisting of those members of $\mathscr{P}$ having degree at most m.

The following elegant result surely deserves to be better known. It has its origins in the 1917 paper [**7**] of Ernst Fischer; see the discussion of (2.8) on page 459 of [**10**]. Let's call it

Fischer's Lemma. *For $f \in \mathscr{P}$ define $L(f) = \Delta(fb)$. Then L is a linear, degree-preserving, bijection of $\mathscr{P}$ onto itself.*

Proof: Clearly L is a linear operator on $\mathscr{P}$. Suppose that $L(f) = 0$ for some $f \in \mathscr{P}$. Let $u = fb$ so that $\Delta u(x) = 0$ for all $x \in \mathbb{R}^n$ and $u(x) = 0$ for all $x \in \partial\mathscr{E}$. By Corollary 2, $u(x) = 0$ for all $x \in \overline{\mathscr{E}}$ so $f(x) = 0$ for all $x \in \mathscr{E}$ since $b(x) > 0$ for all $x \in \mathscr{E}$. But f is a polynomial; hence $f(x) = 0$ for all $x \in \mathbb{R}^n$. We have shown that L is one-to-one.

Now suppose that $0 \le m \in \mathbb{Z}$. If $f \in \mathscr{P}_m$ then $fb \in \mathscr{P}_{m+2}$ and hence $\Delta(fb) \in \mathscr{P}_m$. That is, L maps $\mathscr{P}_m$ *into* itself. Since $\mathscr{P}_m$ is finite dimensional and L is linear and one-to-one, L maps $\mathscr{P}_m$ *onto* itself. ∎

2. THE DIRICHLET PROBLEM FOR ELLIPSOIDS AND POLYNOMIAL BOUNDARY DATA.

Fischer's Lemma and the Maximum Principle, yield a simple proof of

Theorem 1. *Suppose that $f \in \mathscr{P}_m$ for some $m \ge 0$. Then there exists a unique u in $\mathscr{P}_m$ such that*

$$\Delta u(x) = 0 \quad \textit{for all } x \in \mathbb{R}^n \quad \textit{and} \quad u(x) = f(x) \quad \textit{for all } x \in \partial\mathscr{E}. \qquad (\#)$$

Proof: If $m \leq 1$ the conclusion holds with $u = f$. Suppose that $m \geq 2$ and the degree of f is at least 2. We look for a u of the form $f + vb$ with $v \in \mathscr{P}_{m-2}$. For any such u, $u(x) = f(x)$ for all $x \in \partial\mathscr{E}$ and $\Delta u = \Delta f + \Delta(vb)$. By Fischer's Lemma, there exists a unique g in $\mathscr{P}_{m-2}$ such that $\Delta(gb) = -\Delta f$. Thus, if we define $u := f + gb$, then $u \in \mathscr{P}_m$ and (#) holds. Uniqueness follows from Corollary 2. ∎

3. THE MEAN–VALUE PROPERTY AND THE WEIERSTRASS APPROXIMATION THEOREM. Let $B = \{x \in \mathbb{R}^n : |x| \leq 1\}$, $S = \partial B = \{x \in \mathbb{R}^n : |x| = 1\}$, and $B_r(a) = \{x \in \mathbb{R}^n : |x - a| \leq r\} = \{a + ry : y \in B\}$ for $a \in \mathbb{R}^n$ and $r > 0$. Denote by σ the normalized, $n - 1$ dimensional surface measure on S. The following result is a combination of [**1**, Theorems 1.2 and 1.20], which depend mainly upon the Divergence Theorem for B, a fairly self-contained exposition of which can be found in [**3**].

The Mean–Value Property. *Suppose that Ω is a region in $\mathbb{R}^n$, $u : \Omega \to \mathbb{R}$, and u is continuous. Then u is harmonic on Ω if and only if*

$$u(a) = \int_S u(a + rs)\, d\sigma(s) \quad \textit{whenever } B_r(a) \subset \Omega.$$

According to [**6**, p. 35], this theorem can be traced back to an 1840 paper of Gauss.

Corollary. *If Ω is a region in $\mathbb{R}^n$, $u_k : \Omega \to \mathbb{R}$ is harmonic for each $k \in \mathbb{N}$, and $\{u_k\}_{k=1}^{\infty}$ converges uniformly on Ω to $u : \Omega \to \mathbb{R}$, then u is harmonic on Ω.*

Sketch of Proof. Suppose $B_r(a) \subset \Omega$. Then

$$u_k(a) = \int_S u_k(a + rs)\, d\sigma(s) \quad \text{for } k \in \mathbb{N} \quad \text{and} \quad u(a) = \lim_{k \to \infty} u_k(a).$$

Moreover, since each u_k is continuous and the convergence is uniform on Ω, u is continuous on Ω and

$$\lim_{k \to \infty} \int_S u_k(a + rs)\, d\sigma(s) = \int_S u(a + rs)\, d\sigma(s).$$

Hence

$$u(a) = \int_S u(a + rs)\, d\sigma(s).$$

By the Mean Value Property, u is harmonic on Ω. ∎

It does not appear to be well known that, in [**13**], Weierstrass proved his famous approximation theorem not only in the case of a single real variable but also in higher dimensions. That "approximate identity" proof, in the one dimensional case, is the subject of Chapter 59 of the beautiful book of Körner [**11**]. The same proof extends to higher dimensions without serious difficulty, see [**9**, p. 209 and Problem 1 on p. 213]. Because of its geometric appeal, its intimate relationship with the heat equation, and the fact that it affords C^m approximation, Weierstrass's own proof, in the author's opinion, has not been bettered.

The Weierstrass Approximation Theorem. *Given a rectangle I in $\mathbb{R}^n$, a continuous function $f : I \to \mathbb{R}$, and $\epsilon > 0$, there exists a p in $\mathscr{P}$ such that $|f(x) - p(x)| < \epsilon$ for all $x \in I$.*

By a *rectangle* in $\mathbb{R}^n$ we mean a product of n closed bounded intervals.

4. THE DIRICHLET PROBLEM FOR ELLIPSOIDS; A SOLUTION

Theorem 2. *Given a continuous function* $f : \partial\mathscr{E} \to \mathbb{R}$, *there is a unique continuous function* $u : \overline{\mathscr{E}} \to \mathbb{R}$ *such that* u *is* C^2 *on* $\mathscr{E}$, $\Delta u(x) = 0$ *for all* $x \in \mathscr{E}$, *and* $u(x) = f(x)$ *for all* $x \in \partial\mathscr{E}$.

Proof: For $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ let $\nu(x) = \sum_{k=1}^n x_k^2/r_k^2$ and

$$\tilde{f}(x) = \begin{cases} \nu(x) f\big(\nu(x)^{-1} x\big) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

Then $\tilde{f}$ is continuous on $\mathbb{R}^n$ and therefore, by Weierstrass's Theorem, can be uniformly approximated on each rectangle by a polynomial. But $\tilde{f}(x) = f(x)$ for all $x \in \partial\mathscr{E}$. Hence there is a sequence $\{f_k\}_{k=1}^\infty$ in $\mathscr{P}$ that converges to f uniformly on $\partial\mathscr{E}$.

According to Theorem 1, for each $k \in \mathbb{N}$ there is a unique $u_k \in \mathscr{P}$ such that $\Delta u_k(x) = 0$ for all $x \in \mathbb{R}^n$ and $u_k(x) = f_k(x)$ for all $x \in \partial\mathscr{E}$. By the Maximum Principle, for $j, k \in \mathbb{N}$ and $x \in \overline{\mathscr{E}}$,

$$\left|u_j(x) - u_k(x)\right| \le \max\left\{\left|f_j(y) - f_k(y)\right| : y \in \partial\mathscr{E}\right\} \to 0 \quad \text{as } j, k \to \infty.$$

Thus $\{u_k\}_{k=1}^\infty$ converges uniformly on $\overline{\mathscr{E}}$ to a continuous function $u : \overline{\mathscr{E}} \to \mathbb{R}$.

By the Corollary to the Mean Value Property, u is harmonic on $\mathscr{E}$. Moreover, for $x \in \partial\mathscr{E}$, $u(x) = \lim_{k\to\infty} u_k(x) = \lim_{k\to\infty} f_k(x) = f(x)$. ∎

5. A FINITE DIMENSIONAL DIRICHLET PRINCIPLE. Inspired by Gårding's discussion of the Dirichlet principle, [**8**, pp. 96–98], we present

Another Proof of Theorem 1. Assume $m \ge 2$ and $f \in \mathscr{P}_m$. For $\psi \in \mathscr{P}_m$ and $x \in \mathbb{R}^n$ let $\nabla\psi(x) = (\partial_1\psi(x), \dots, \partial_n\psi(x))$, the *gradient* of ψ at x, and let

$$D(\psi) = \int_{\mathscr{E}} |\nabla\psi(x)|^2\, dx \text{—the } \textit{Dirichlet integral}.$$

For $\psi, \chi \in \mathscr{P}_m$ define $B(\psi, \chi) = \int_{\mathscr{E}} \nabla\psi(x) \cdot \nabla\chi(x)\, dx$. Notice that B is a symmetric bilinear form on $\mathscr{P}_m$ and $B(\psi, \psi) = D(\psi) \ge 0$ for all $\psi \in \mathscr{P}_m$. Moreover, for $\psi \in \mathscr{P}_m$, $D(\psi) = 0$ if and only if ψ is constant (i.e., $\psi \in \mathscr{P}_0$).

Let $V = \{wb : w \in \mathscr{P}_{m-2}\}$ and $A = \{f + v : v \in V\}$. Note that V is a linear subspace of $\mathscr{P}_m$ and every member of V vanishes on $\partial\mathscr{E}$. Hence, if we let $\langle \psi, \chi \rangle = B(\psi, \chi)$ for $\psi, \chi \in V$, then $\langle \cdot, \cdot \rangle$ is an inner product for V and its associated norm satisfies $\|v\|^2 = B(v, v) = D(v)$ for $v \in$ V.

Observe that A is an affine subspace of $\mathscr{P}_m$ and if $u \in A$ then $u(x) = f(x)$ for all $x \in \partial\mathscr{E}$.

We aim to prove that D has a unique minimizer on A, say u, and this u is a solution to our problem. For any $v \in V$ we have

$$D(f + v) = \int_{\mathscr{E}} |\nabla f(x) + \nabla v(x)|^2\, dx = D(f) + 2B(f, v) + \|v\|^2.$$

Now the map $v \mapsto B(f, v)$, for $v \in V$, is a linear functional on the finite dimensional vector space V. Hence there is a unique g in V such that $B(f, v) = \langle g, v \rangle$ for all $v \in V$. For $v \in V$ we therefore have

$$D(f + v) = D(f) + 2\langle g, v \rangle + \|v\|^2 = D(f) + \|v + g\|^2 - \|g\|^2$$

and this is clearly least exactly when $v = -g$. Let $u = f - g \in A$ and conclude that $D(u) \le D(\psi)$ for all $\psi \in A$.

Suppose that $0 \neq v \in V$. If $t \in \mathbb{R}$ then $u + tv \in A$ so that $D(u) \leq D(u + tv) = D(u) + 2tB(u, v) + t^2\|v\|^2$, i.e., $0 \leq 2tB(u, v) + t^2\|v\|^2$ for all $t \in \mathbb{R}$. It follows that, for all $v \in V$, $0 = B(u, v) = \sum_{k=1}^{n} \int_{\mathscr{E}} \partial_k u(x) \partial_k v(x)\, dx$. But recall that every member of V vanishes on $\partial\mathscr{E}$. Integration by parts therefore leads us to conclude that $0 = \sum_{k=1}^{n} \int_{\mathscr{E}} (\partial_k^2 u(x)) v(x) dx = \int_{\mathscr{E}} \Delta u(x) v(x)\, dx$ for all $v \in V$. That is,

$$0 = \int_{\mathscr{E}} \Delta u(x) w(x) b(x)\, dx \quad \text{for every } w \in \mathscr{P}_{m-2}. \qquad (*)$$

Now $\Delta u \in \mathscr{P}_{m-2}$ and the map $(\psi, \chi) \mapsto \int_{\mathscr{E}} \psi(x)\chi(x) b(x)\, dx$ $(\psi, \chi \in \mathscr{P}_{m-2})$ is clearly an inner product for $\mathscr{P}_{m-2}$. Hence, $(*)$ ensures that $\Delta u = 0$. Since $u \in A$, $u(x) = f(x)$ for $x \in \partial\mathscr{E}$ and uniqueness is guaranteed by Corollary 2 of the Maximum Principle. ■

6. REMARKS

(i) Theorem 1 was proved in yet another nonstandard way in [**5**, Théorme 6, p. 60].

(ii) Suppose that $g \in \mathscr{P}$ and we are interested in solving the Dirichlet problem for *Poisson's equation*: $\Delta u(x) = g(x)$, $x \in \mathscr{E}$, and $u(x) = f(x)$, $x \in \partial\mathscr{E}$ where $f : \partial\mathscr{E} \to \mathbb{R}$ is a given continuous function. By Fischer's Lemma there exists a v in $\mathscr{P}$ such that $g = \Delta(vb)$. By Theorem 2, there exists a continuous $w : \overline{\mathscr{E}} \to \mathbb{R}$ such that w is C^2 on $\mathscr{E}$, $\Delta w(x) = 0$, for $x \in \mathscr{E}$, and $w(x) = f(x)$ for $x \in \partial\mathscr{E}$. Let $u(x) = w(x) + v(x)b(x)$ for $x \in \overline{\mathscr{E}}$. Then u is continuous on $\overline{\mathscr{E}}$ and C^2 on $\mathscr{E}$, $\Delta u(x) = \Delta w(x) + \Delta(vb)(x) = g(x)$ for $x \in \mathscr{E}$, and $u(x) = w(x) = f(x)$ for $x \in \partial\mathscr{E}$.

(iii) Theorems 1 and 2 apply to arbitrary ellipsoids and not just the "canonical" types considered so far. To see this it suffices to note that every ellipsoid is isometric to one of the kind we've considered and to check that the class of harmonic function is invariant under isometries; [**1**, pp. 2–3].

(iv) In place of the Laplacian one could substitute an operator of the form $\sum_{k=1}^{n} \lambda_k \partial_k^2$ with positive real λ_k.

(v) Extensions of Fischer's ideas and applications thereof to differential problems have been given by several authors; see [**10**], [**12**], and the references included therein.

(vi) Additional intriguing properties of harmonic polynomials, together with applications thereof to boundary value problems for B, can be found in the charming paper of Axler and Ramey [**2**].

ACKNOWLEDGMENTS. I am grateful to my colleague David Siegel for bringing Fischer's Lemma, [**5**], and [**10**] to my attention. I thank Dmitry Khavinson, University of Arkansas, for valuable correspondence. This work was supported by NSERC (Canada) Grant #7153.

REFERENCES

1. S. Axler, P. Bourdon, and W. Ramey, *Harmonic Function Theory*, Springer-Verlag, New York, 1992.
2. S. Axler and W. Ramey, Harmonic Polynomials and Dirichlet-type Problems, *Proc. Amer. Math. Soc.* **123** (1995) 3765–3773.
3. John A. Baker, Integration Over Spheres and the Divergence Theorem for Balls, *Amer. Math. Monthly* **104** (1997) 36–47.
4. Umberto Bottazzini, *The Higher Calculus: A History of Real and Complex Analysis from Euler to Weierstrass*, Springer-Verlag, New York, 1986.
5. Marcel Brelot and Gustave Choquet, Polynômes harmoniques et polyharmoniques, *Colloque sur les équations aux dérivées partielle*, Brussels (1954) 45–66.

6. J. Dieudonné, *History of Functional Analysis*, North-Holland, Amsterdam, 1981.
7. E. Fischer, Über die Differentiationsprozesse der Algebra, *J. für Math.* **148** (1917) 1–78.
8. Lars Gårding, *Encounter with Mathematics*, Springer-Verlag, New York, 1977.
9. Fritz John, *Partial Differential Equations, Fourth Edition*, Springer-Verlag, New York, 1982.
10. Dmitry Khavinson and Harold S. Shapiro, Dirichlet's problem when the data is an entire function, *Bull. London Math. Soc.* **24** (1992) 456–468.
11. T. W. Körner, *Fourier Analysis*, Cambridge University Press, Cambridge, 1989.
12. Harold S. Shapiro, An Algebraic Theorem of E. Fischer and the Holomorphic Goursat Problem, *Bull. London Math. Soc.* **21** (1989) 513–537.
13. Karl Weierstrass, *Math Werke*, Bd. III, Berlin: Mayer u. Muller, 1903, 1–37.

JOHN BAKER, since taking early retirement in 1996 and becoming (undistinguished) Professor Emeritus, has continued to engage in research and expository writing. He continues to occupy the bucket seats of classical analysis and functional equations at the University of Waterloo. A major change in his nonacademic activities involves a refocusing of his curmudgeonly writings from university politics to the local variety. In all likelihood, as you read this he'll still be involved in a battle to keep a box store development from desecrating the rural community in which he lives.
Department of Pure Mathematics, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada
jabaker@math.uwaterloo.ca

Said a mathematician named Haar,
"Von Neumann can't see very far.
He missed a great treasure:
They call it Haar measure.
Poor Johnny's just not up to par."

Contributed by Paul R. Chernoff, UC Berkeley, who provides the following background: John von Neumann had proved the existence of invariant measures on compact topological groups, but tried to discourage Alfred Haar from working on the locally compact case on the grounds that it seemed unlikely to be true in that generality. Fortunately, Haar persisted.

Curves Whose Curvature Depends on Distance From the Origin

David A. Singer

1. INTRODUCTION. The fundamental existence and uniqueness theorem for curves in Euclidean space $\mathbb{R}^3$ states that a curve is uniquely determined up to rigid motion by its curvature and torsion, given as functions of its arc-length. Furthermore, given continuous functions $\kappa(s)$ and $\tau(s)$, with $\kappa(s)$ positive and continuously differentiable, there is a differentiable curve (of class at least C^3) with curvature κ and torsion τ.

In practice, such curves are often impossible to find explicitly, due to the difficulty in solving the *Frenet Equations*, the linear differential equations governing the curve; but see [**2**] for an example where the equations can be solved. In general, a result of Lie and Darboux shows that solving these equations is equivalent to solving a certain complex Riccati equation; see [**3**, p. 36] for details. Happily, in the planar case the Frenet equations can always be integrated by quadratures.

We consider a different sort of problem. Suppose the curvature of a proposed curve in the plane is given in terms of its *position*. Can the curve be determined, and if so, how? The general form of this problem requires one to solve a *nonlinear* differential equation:

$$\frac{x'(t)y''(t) - y'(t)x''(t)}{\left(x'(t)^2 + y'(t)^2\right)^{\frac{3}{2}}} = \kappa(x(t), y(t)) \tag{1.1}$$

An interesting solved example of this problem occurs when the curvature is proportional to one of the coordinate functions, say $\kappa(x, y) = cy$. This is a remarkable property of the *Euler elasticae*, curves that minimize $\int \kappa(s)^2\, ds$ among curves of fixed length with fixed first-order boundary conditions. These are the "natural splines," formed by taking a thin inextensible wire of uniform thickness and pinning and "welding" the two ends in fixed positions; see [**4**].

Among the Euler elasticae there is a unique *closed* curve, which is in the shape of a figure-eight. The curvature of this curve is given by the elliptic function $\operatorname{sn}(u, p)$, where u is proportional to the arc-length parameter and $p = 0.9089086\ldots$, the elliptic modulus, satisfies the transcendental equation $2E(p) = K(p)$. Here K and E are the complete elliptic integrals of the first and second kind. The curvature vanishes at the crossing point, and the x-axis bisects the figure, with one loop above and the other below the axis.

2. A FIRST ATTEMPT. The curvature of a planar curve is most simply defined using a coordinate system that moves with the curve. Let T be the unit tangent vector to the curve $X(s)$, and let N be the vector orthogonal to T such that the

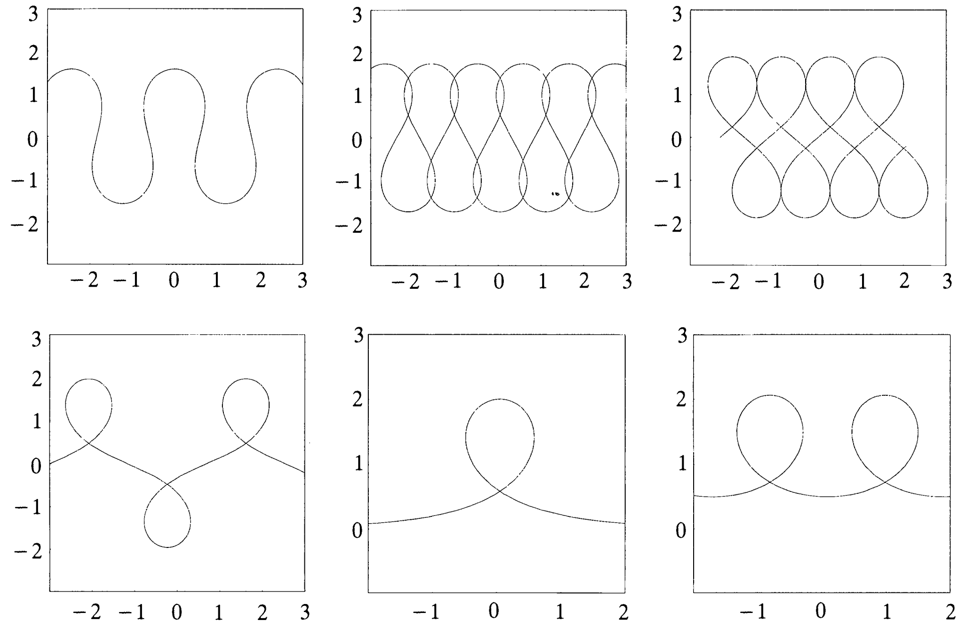


Figure 1. Some Euler elasticae.

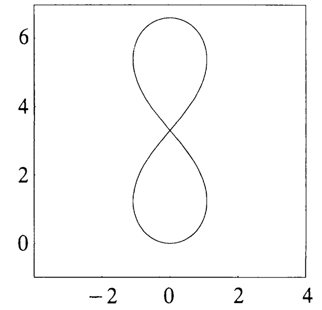


Figure 2. The figure-eight Euler elastica.

frame T, N is positively oriented. If s is the arc-length parameter, then $dX/ds = T$, and the Frenet equations are

$$\frac{dT}{ds} = \kappa N, \quad \frac{dN}{ds} = -\kappa T. \tag{2.1}$$

If we define the angle θ by writing the Euclidean coordinates of T as $(\cos\theta, \sin\theta)$, then we can rewrite (2.1) as

$$\frac{d\theta}{ds} = \kappa(s), \quad \frac{dx}{ds} = \cos\theta, \quad \text{and} \quad \frac{dy}{ds} = \sin\theta. \tag{2.2}$$

Equations (2.2) show that once κ has been determined, the curve can be found by three quadratures. They are also very useful in computing numerical and graphical solutions to the Frenet equations in the plane.

The condition we seek to have the curve satisfy is $\|X\| = \kappa$. A direct assault on (2.2) after substituting $\kappa = \sqrt{x^2 + y^2}$ is unpromising. A more straightforward

approach might be to write $X(s) = (\kappa(s)\cos\phi, \kappa(s)\sin\phi)$ in polar coordinates. This leads to a pair of coupled second-order equations, namely

$$\frac{d^2\kappa}{ds^2} = \kappa\left(\frac{d\phi}{ds}\right)^2 - \kappa^2\frac{d\phi}{ds}, \quad 2\frac{d\kappa}{ds}\frac{d\phi}{ds} + \kappa\frac{d^2\phi}{ds^2} = \kappa\frac{d\kappa}{ds}, \tag{2.3}$$

which can be reduced by use of the first integral

$$\left(\frac{d\kappa}{ds}\right)^2 + \kappa^2\left(\frac{d\phi}{ds}\right)^2 = 1 \tag{2.4}$$

to a (somewhat unpleasant) second-order differential equation for κ. Once this is solved the curve can be determined by quadratures.

A slightly more devious approach is to define functions $\alpha(s)$ and $\beta(s)$ by the formula

$$X = \alpha T + \beta N \tag{2.5}$$

in order to parametrize the curve by coordinates that move with the curve. Differentiating (2.5) gives

$$\frac{d\alpha}{ds} = \beta\kappa + 1, \quad \frac{d\beta}{ds} = -\alpha\kappa. \tag{2.6}$$

The condition $\|X\| = \kappa$ in these coordinates is

$$\alpha^2 + \beta^2 = \kappa^2, \tag{2.7}$$

so we may define an angle ψ by $\alpha = \kappa\cos\psi$, $\beta = \kappa\sin\psi$. The equations (2.6) now lead to a pair of first order equations:

$$\frac{d\kappa}{ds} = \cos\psi, \quad \kappa\frac{d\psi}{ds} + \kappa^2 = -\sin\psi. \tag{2.8}$$

Unfortunately, it is again not clear how to solve the differential equations and determine κ.

3. MOTION UNDER GRAVITATIONAL FORCE. Although it is not always possible to find explicit solutions to a system of second-order differential equations, this can be done by quadratures for completely integrable Hamiltonian systems. Thus, we now attempt to reformulate our problem as a Hamiltonian system with sufficient symmetry to be integrable.

A fundamental example of such a system arises from the problem of an orbiting object. Newton's equations of motion are $\vec{F}(X) = mX''$, where X is the position of the object. Suppose the object is moving through space under the influence of a central force $\vec{F}(X) = -f(r)X$, where $r = \|X\|$ is the distance from the origin and $f(r)$ is some continuous function. The motion satisfies the second order differential equations

$$X'' = -\frac{1}{m}f(r)X, \tag{3.1}$$

and the motion satisfies Kepler's Second Law (conservation of angular momentum): $X \times X' = C$, a constant vector, as can easily be seen by differentiating. The motion lies in a plane, which we assume is the (x, y)-plane, and if we put the curve in polar coordinates $X = (r\cos\phi, r\sin\phi)$, then

$$r^2\frac{d\phi}{dt} = \|X \times X'\| = c, \tag{3.2}$$

where $c = \|C\|$ is a constant (the angular momentum).

Now define a function $\Phi(r)$ by

$$\frac{d\Phi}{dr} = rf(r).$$

Then $\vec{F}(X) = -\nabla V$, where $V(x, y, z) = \Phi(\sqrt{x^2 + y^2 + z^2})$. The motion satisfies conservation of energy:

$$\frac{1}{2}m\|X'\|^2 + V = \frac{1}{2}m\left(\left(\frac{dr}{dt}\right)^2 + r^2\left(\frac{d\phi}{dt}\right)^2\right) + \Phi(r) = E_0, \tag{3.3}$$

where E_0 is a constant (the total energy).

Using (3.2) and (3.3), it is possible to solve the equations of motion by quadrature. In the special case where $\Phi(r) = -1/r$, the force satisfies the inverse square law and we have the Kepler problem. The solutions in that case are ellipses, parabolas, and hyperbolas.

Now let us compute the curvature of solutions. Since the parametrization is no longer by arc length, we need the general formula for the curvature of a curve:

$$\kappa(t) = \frac{\|X' \times X''\|}{\|X'\|^3}. \tag{3.4}$$

Plugging in (3.1) and (3.2), this equation becomes

$$\kappa(t) = \frac{cf(r)}{m\|X'\|^3}. \tag{3.5}$$

Now using (3.3) to compute the denominator, we arrive at the formula

$$\kappa(t) = \frac{c\sqrt{m}\Phi'(r)}{\sqrt{8}r(E_0 - \Phi(r))^{\frac{3}{2}}}. \tag{3.6}$$

Note that the value of κ depends only on r. Define $\mu(r) = (E_0 - \Phi(r))^{-\frac{1}{2}}$. Then (3.6) reduces to

$$\frac{d\mu}{dr} = \frac{\sqrt{2}}{c\sqrt{m}}r\kappa(r). \tag{3.7}$$

If $\kappa(r)$ is any function such that $r\kappa(r)$ is continuous, we can solve (3.7) for μ. Define $\Phi(r) = -1/\mu(r)^2$, and consider solutions to (3.1) with energy $E_0 = 0$. Note that a specific choice of μ also specifies a value of the momentum c_0. That is, a solution to (3.1) has curvature *proportional* to the given function provided it has energy 0; its curvature is equal to the given function if its angular momentum $c = c_0$. This is part of the proof of our main result:

Theorem 3.1. *Let $\kappa(r)$ be a function such that $r\kappa(r)$ is continuous. Then the problem of determining a curve whose curvature is $\kappa(r)$, where r is the distance from the origin, is solvable by quadratures.*

4. THE CASE $\kappa(r) = r$. Using (2.2), it is not difficult to produce pictures of curves whose curvature is equal to the distance from the origin. Some of them are illustrated in Figure 3. We want to find analytic representations for these curves.

Theorem 3.1 is slightly deceptive. The fact that the differential equation is integrable by quadratures does not mean that it is easy to perform the integrations, as we now illustrate with the case that inspired this paper: $\kappa(r) = r$. The first step,

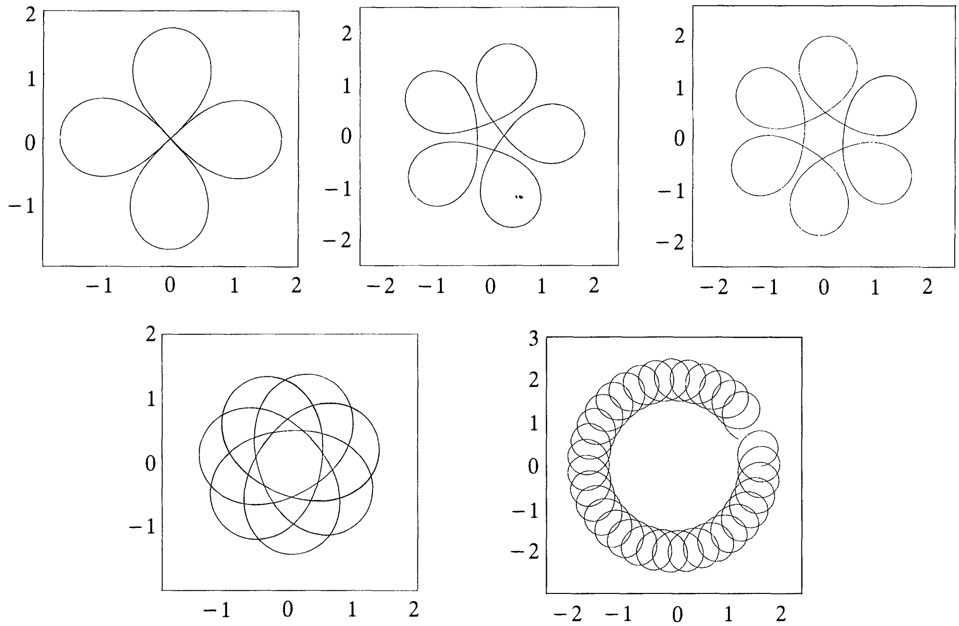


Figure 3. Curves with curvature equal to radial distance.

solving (3.7) is easy and gives

$$\mu(r) = \frac{\sqrt{2}}{3c_0\sqrt{m}}(r^3 + a), \quad \Phi(r) = -\frac{9c_0^2 m}{2(r^3 + a)^2}, \tag{4.1}$$

where a is a constant of integration. This corresponds to a force law in which the magnitude is proportional to $r^2/(r^3 + a)^3$.

Applying (3.3) and assuming that $r^2\phi' = c_0$, we obtain

$$(r')^2 + r^2(\phi')^2 = \frac{9c_0^2}{(r^3 + a)^2} = \frac{9r^4(\phi')^2}{(r^3 + a)^2}. \tag{4.2}$$

This has among its solutions the circular orbit $X(t) = (r_0 \cos(\alpha t), r_0 \sin(\alpha t))$, where $\alpha = 3c_0/r_0(r_0^3 + a)$ and $r_0^2\alpha = c_0$. This implies $3r_0 = r_0^3 + a$. However, X is a solution to (3.1) only when $\alpha^2 = 27c_0^2 r_0/(r_0^3 + a)^3$. This implies $r_0 = 1$, which of course we knew! Other solutions can have $r' = 0$ only at isolated points.

We can eliminate t and find $\phi = \phi(r)$ from

$$\phi = \pm \int \frac{r^3 + a}{r\sqrt{9r^2 - (r^3 + a)^2}}\, dr. \tag{4.3}$$

This is not always an elementary integral. One special case, however, is very pleasant, namely the case where $a = 0$. Then the integral becomes elementary, and the resulting curve is given by

$$\phi - \phi_0 = \frac{1}{2}\arcsin\left(\frac{r^2}{3}\right), \quad r^2 = 3\sin 2(\phi - \phi_0). \tag{4.4}$$

This curve is none other than (one leaf of) the *Bernoulli lemniscate* shown in Figure 4.

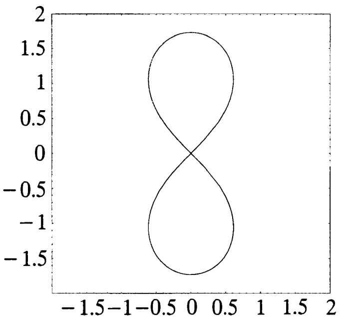


Figure 4. The Bernoulli lemniscate.

From this example it becomes clear why the previous, more straightforward, attempts led to such complicated equations. The solution to the original equations would have given the lemniscate parametrized by arc length. However, the arc length function for the lemniscate is an elliptic integral. Thus even this special case would be much more complicated to find. In fact, we have not yet solved (3.1) explicitly, even in the special case; we found instead a non-parametric representation of the trajectory. This is not difficult to do, however, as we observe in the next section.

5. SUB-AFFINE ARC LENGTH. Solutions to equations of motion for particles moving under the influence of a central force have a natural parametrization. From the equation $X'' = g(x, y)X$ we have observed that $\|X \times X'\|$ is constant. Now let $X(t)$ be a curve satisfying $X \times X' \neq \vec{0}$. Then there is a re-parametrization of the curve such that with respect to the new parameter $\sigma = \sigma(t)$ the curve satisfies

$$\left\| X(\sigma) \times \frac{dX}{d\sigma} \right\| \equiv 1. \tag{5.1}$$

This parameter is computed by

$$\sigma(t) = \int_0^t \|X \times X'\| \, dt. \tag{5.2}$$

In the case of the lemniscate $r^2 = 3\cos(2\theta)$, evaluation of (5.2) is an elementary calculation and yields $\sigma = \frac{3}{2}\sin(2\theta)$.

The parameter σ may be called the *sub-affine arc length parameter*. The motivation for this name is the following: If $Y(t)$ is a curve in the plane such that the vectors $Y'(t)$ and $Y''(t)$ are linearly independent, then we may re-parametrize the curve by a parameter σ such that with respect to this new parametrization $\|Y' \times Y''\| \equiv 1$. This parametrization is known as the *affine arc length*; see [**1**, p. 149]. It plays a role in affine geometry exactly analogous to the usual arc length parameter in Euclidean geometry. If a curve is parametrized by affine arc-length, then $Y''' + \rho Y' = 0$, where $\rho(\sigma)$ is the *affine curvature* of the curve Y; it is a geometric invariant of the curve under unimodular affine transformations of the plane.

If Y is a curve parametrized by affine arc length, then its derivative $X(\sigma) = dY/d\sigma$ is parametrized by subaffine arc length. We call $\rho(\sigma) = -d^2X/d\sigma^2$ the *subaffine curvature* of X. This is a geometric invariant of X under unimodular linear transformations of R^2.

Our problem was to find curves whose curvature is equal to the distance from the origin. Let γ be such a curve. If $\gamma(t) \times \gamma'(t) \neq 0$, then we can locally parametrize γ by subaffine arc length. Then

$$\frac{d^2\gamma}{d\sigma^2} = -\rho(\sigma)\gamma(\sigma). \tag{5.3}$$

The Euclidean curvature is given by the formula $\kappa(\gamma(\sigma)) = \rho(\sigma)/\|\gamma'(\sigma)\|^3$.

If we now assume that the radial distance from the origin varies monotonically on some interval $\sigma_0 \leq \sigma \leq \sigma_1$, we can solve for σ in terms of r and define $f(r) = \rho(\sigma(r))$. Thus, on any part of a curve along which $X \times X'$ does not vanish and $X \cdot X'$ does not vanish, the curve arises as a solution to an equation of the form (3.1).

Local extrema of r do not present any difficulty, since it is evident that the curve has a symmetry at each such point. Places where $X \times X' = 0$, however, represent limits of trajectories. For example, in the example of the lemniscate the origin is a singularity of the orbit. Note, however, that these are of necessity isolated points on the curve. This observation completes the proof of the theorem.

It is interesting to note that the graphical solution corresponding to the lemniscate actually produces *two* lemniscates at right angles to each other. Indeed, the solution to (2.2) with initial conditions $x = 0, y = 0$, θ arbitrary, sweeps out alternate halves of the two lemniscates, producing a flower with four loops. Note that the equation we have solved is for the *signed* curvature equaling the radial distance. Thus κ is required to remain non-negative. The curvature of the lemniscate changes sign as it passes through the origin.

ACKNOWLEDGMENT. This paper arose when Prof. Oscar Garay of Universidad de Granada asked the author if he could find a curve whose curvature is proportional to the distance from a single point rather than a line. The author is grateful to him for many helpful and interesting discussions.

REFERENCES

1. H. W. Guggenheimer, *Differential geometry*, Dover, New York, 1977.
2. P. D. Scofield, Curves of constant precession, *Amer. Math. Monthly* **102** (1995) 531–537.
3. D. J. Struik, *Differential geometry*, second ed., Addison-Wesley, Reading, Massachusetts, 1961.
4. C. Truesdell, The influence of elasticity on analysis: The classic heritage, *Bull. Amer. Math. Soc.* **9** (1983) 293–310.

DAVID SINGER learned mathematics and how to teach from his mother and at the University of Pennsylvania, receiving his Ph.D. under Herman Gluck in 1970. After a National Science Foundation postdoctoral fellowship at Princeton and teaching at Cornell University, he came to Case Western Reserve University in Cleveland in 1975, where he has pursued his research interests in differential geometry and Hamiltonian and other dynamical systems. He is also very interested in issues of mathematics education at K–12 schools and has recently written a geometry text for teachers.
Case Western Reserve University, Cleveland, Ohio 44106-7058
das5@po.cwru.edu

The Fifty-Ninth William Lowell Putnam Mathematical Competition

Leonard F. Klosinski, Gerald L. Alexanderson, and Loren C. Larson

The results of the Fifty-ninth William Lowell Putnam Mathematical Competition held December 5, 1998, follow. They have been determined in accordance with the regulations governing the competition, a contest supported by the William Lowell Putnam Prize Fund for the Promotion of Scholarship, a fund endowed by Mrs. Putnam in memory of her husband. The annual Competition is held under the auspices of the Mathematical Association of America.

The first prize, \$25,000, was awarded to the Department of Mathematics at Harvard University. The members of the winning team were Michael L. Develin, Ciprian Manolescu, and Dragos N. Oprea; each was awarded a prize of \$1000.

The second prize, \$20,000, was awarded to the Department of Mathematics at the Massachusetts Institute of Technology. The members of the winning team were Amit Khetan, Eric H. Kuo, and Edward D. Lee; each was awarded a prize of \$800.

The third prize, \$15,000, was awarded to the Department of Mathematics at Princeton University. The members of the winning team were Craig R. Helfgott, Michael R. Korn, and Yuliy V. Sannikov; each was awarded a prize of \$600.

The fourth prize, \$10,000, was awarded to the Department of Mathematics at the California Institute of Technology. The members of the winning team were Christopher C. Chang, Christopher M. Hirata, and Hanhui Yuan; each was awarded a prize of \$400.

The fifth prize, \$5,000, was awarded to the Department of Mathematics at the University of Waterloo. The members of the winning team were Sabin Cautis, Derek I. E. Kisman, and Soroosh Yazdani; each was awarded a prize of \$200.

The five highest ranking individual contestants, in alphabetical order, were Nathan G. Curtis, Duke University; Michael L. Develin, Harvard University; Kevin D. Lacker, Duke University; Ciprian Manolescu, Harvard University; and Ari M. Turner, Princeton University. Each of these has been designated a Putnam Fellow by the Mathematical Association of America and awarded a prize of \$2,500 by the Putnam Prize Fund.

The next ten highest ranking contestants, in alphabetical order, were: Sabin Cautis, University of Waterloo; Christopher C. Chang, California Institute of Technology; Donny C. Cheung, University of Waterloo; Dragos F. Ghioca, University of Pittsburgh; Andrei C. Gnepp, Harvard University; Michael Ostrovsky, Stanford University; Colin A. Percival, Simon Fraser University; Alexander H. Saltman, Harvard University; Daniel A. Stronger, Harvard University; and Liang Yang, Yale University; each was awarded a prize of \$1,000.

The next eleven highest ranking contestants, in alphabetical order, were: Ian W. Caines, Dalhousie University; Scott H. Carnahan, California Institute of Technology; Adrian D. Corduneanu, University of Toronto; Matthew T. Gealy, University of Chicago; Larry D. Guth, Yale University; Amit Khetan, Massachusetts Institute of Technology; Michael R. Korn, Princeton University; Abhinav Kumar, Massachusetts Institute of Technology; Davesh Maulik, Harvard University;

Dragos Nicolae Oprea, Harvard University; and Ronfeng Sun, Clark University; each was awarded a prize of $250.

The following teams, named in alphabetical order, received honorable mention: University of Chicago, with team members Benjamin M. Cowan, Matthew T. Gealy, and Christopher D. Malon; Duke University, with team members Nathan G. Curtis, Andrew O. Dittmer, and Carl A. Miller; The Johns Hopkins University, with team members Alexander J. Diesl, Rakesh M. Lal, and Nehal S. Munshi; Stanford University, with team members Eugene G. Davydov, Alexander S. Dugas, and Michael Ostrovsky; and University of Toronto, with team members Cyrus Chen Hsia, Bhaskara M. Marthi, and Ryan O'Donnell.

Honorable mention was achieved by the following thirty-three individuals named in alphabetical order: Chetan Tukaram Balwe, University of Michigan, Ann Arbor; Adrian Birka, Massachusetts Institute of Technology; Dmitriy S. Boyarchenko, University of Pennsylvania; Li-Chung Chen, Harvard University; Constantin S. Chiscanu, Massachusetts Institute of Technology; John J. Clyde, Duke University; Shai M. Cohen, University of Toronto; Benjamin M. Cowan, University of Chicago; Samit Dasgupta, Harvard University; Kenneth K. Easwaran, Stanford University; Frederik H. Eaton, California Institute of Technology; Brad A. Friedman, University of Illinois, Urbana-Champaign; Craig R. Helfgott, Princeton University; Christopher M. Hirata, California Institute of Technology; Cyrus Chen Hsia, University of Toronto; Kai Huang, Massachusetts Institute of Technology; Liviu Ignat, University of Pittsburgh; Miro Jurisic, Massachusetts Institute of Technology; Scott P. Kempen, Marquette University; Derik I. E. Kisman, University of Waterloo; Eric H. Kuo, Massachusetts Institute of Technology; Chin-Aik Lee, Lebanon Valley College; Edward D. Lee, Massachusetts Institute of Technology; Christopher C. Mihelich, Harvard University; Carl A. Miller, Duke University; Mintcho P. Petkov, Dartmouth College; Dmitry L. Sagalovskiy, Harvard University; Yuliy V. Sannikov, Princeton University; Jan K. Siwanowicz, City College of New York; Mark J. Tilford, California Institute of Technology; Ian W. T. Vander Burgh, University of Waterloo; Cristian Voicu, St. Lawrence University; and Hoeteck Wee, Massachusetts Institute of Technology.

The other individuals who achieved ranks among the top 103, in alphabetical, order of their schools were: University of British Columbia, Lawrence Tang; California Institute of Technology, Ryan L. McCorvie, Michael A. Shulman, Hanhui Yuan, and Kaiwen Xu; Case Western Reserve University, Andrew D. Frohmader; University of Central Florida, Daniel E. Moraseski; University of Chicago, Charles D. Cadmon, Christopher D. Malon, and Sergey Vassiliev; University of Colorado, Boulder, Tao He; Duke University, Andrew O. Dittmer; Georgia Institute of Technology, Jeffrey M. Fowler; University of Georgia, Charles Rollin Mathis; Harvard University, Lukasz Fidkowski, Robert Ribciuc, and Joshua S. Vonkorff; Harvey Mudd College, Ranjithkumar Rajagopalan; The Johns Hopkins University, Alexander J. Diesl; Massachusetts Institute of Technology, Michael L. Brasher, Pokman Cheung, Shamik Das, Ashish Mishra, Brent J. Yen, and Boris Zbarsky; McGill University, Alexandru E. Ghitza; University of Michigan, Ann Arbor, Kurt A. Steinkraus; University of Nebraska, Lincoln, Travis W. Fisher; New York University, Ioana Dumitriu; Northeastern University, Ivo K. Nikolov; University of Pennsylvania, David Futer; Princeton University, Todd W. Geldon; Rice University, Brian David Rothbach; Simon's Rock College, Robert J. Young; University of South Carolina, Jason M. Burns; Stanford University, Eugene V. Davydov, Alex S. Dugas, and Siutaur Pang; University of Toronto, Bhaskara M.

Marthi, and Ryan W. O'Donnell; University of Vermont, Laura P. Riccio; Washington University, St. Louis, Dan B. Johnston; and University of Waterloo, Joel Kamnitzer, and Soroosh Yazdani.

There were 2,581 individual contestants from 419 colleges and universities in Canada and the United States in the Competition of December 5, 1998. Teams were entered by 319 institutions. The Questions Committee for the fifty-ninth Competition consisted of Michael J. Larsen (Chair), University of Pennsylvania; Steven G. Krantz, Washington University, St. Louis; and David J. Wright, Oklahoma State University; they composed the problems and were most prominent among those suggesting solutions. Alternate solutions to some problems have been published in *Mathematics Magazine*.

THE 59th ANNUAL WILLIAM LOWELL PUTNAM EXAMINATION

A1. A right circular cone has base of radius 1 and height 3. A cube is inscribed in the cone so that one face of the cube is contained in the base of the cone. What is the side-length of the cube?

A2. Let s be any arc of the unit circle lying entirely in the first quadrant. Let A be the area of the region lying below s and above the x-axis and let B be the area of the region lying to the right of the y-axis and to the left of s. Prove that $A + B$ depends only on the arc length, and not on the position, of s.

A3. Let f be a real function on the real line with continuous third derivative. Prove that there exists a point a such that $f(a) \cdot f'(a) \cdot f''(a) \cdot f'''(a) \geq 0$.

A4. Let $A_1 = 0$ and $A_2 = 1$. For $n > 2$, the number A_n is defined by concatenating the decimal expansions of A_{n-1} and A_{n-2} from left to right. For example, $A_3 = A_2 A_1 = 10$, $A_4 = A_3 A_2 = 101$, $A_5 = A_4 A_3 = 10110$, and so forth. Determine all n such that 11 divides A_n.

A5. Let $\mathscr{F}$ be a finite collection of open discs in $\mathbf{R}^2$ whose union contains a set $E \subseteq \mathbf{R}^2$. Show that there is a pairwise disjoint subcollection $D_1, \ldots, D_n$ in $\mathscr{F}$ such that

$$\bigcup_{j=1}^{n} 3D_j \supseteq E.$$

Here, if D is the disc of radius r and center P, then $3D$ is the disc of radius $3r$ and center P.

A6. Let A, B, C denote distinct points with integer coordinates in $\mathbf{R}^2$. Prove that if

$$(|AB| + |BC|)^2 < 8 \cdot [ABC] + 1$$

then A, B, C are three vertices of a square. Here $|XY|$ is the length of segment XY and $[ABC]$ is the area of triangle ABC.

B1. Find the minimum value of

$$\frac{(x + 1/x)^6 - (x^6 + 1/x^6) - 2}{(x + 1/x)^3 + (x^3 + 1/x^3)}$$

for $x > 0$.

B2. Given a point (a, b) with $0 < b < a$, determine the minimum perimeter of a triangle with one vertex at (a, b), one on the x-axis, and one on the line $y = x$. You may assume that a triangle of minimum perimeter exists.

B3. Let H be the unit hemisphere $\{(x, y, z): x^2 + y^2 + z^2 = 1,\ z \geq 0\}$, C the unit circle $\{(x, y, 0): x^2 + y^2 = 1\}$, and P the regular pentagon inscribed in C. Determine the surface area of that portion of H lying over the planar region inside P, and write your answer in the form $A \sin \alpha + B \cos \beta$, where A, B, α, and β are real numbers.

B4. Find necessary and sufficient conditions on positive integers m and n so that

$$\sum_{i=0}^{mn-1} (-1)^{\lfloor i/m \rfloor + \lfloor i/n \rfloor} = 0.$$

B5. Let N be the positive integer with 1998 decimal digits, all of them 1; that is,

$$N = \underbrace{1111 \cdots 11}_{1998 \text{ digits}}.$$

Find the thousandth digit after the decimal point of $\sqrt{N}$.

B6. Prove that, for any integers a, b, c, there exists a positive integer n such that $\sqrt{n^3 + an^2 + bn + c}$ is not an integer.

SOLUTIONS. In the 12-tuples $(n_{10}, n_9, n_8, n_7, n_6, n_5, n_4, n_3, n_2, n_1, n_0, n_{-1})$ following each problem number, n_i for $10 \geq i \geq 0$ is the number of students among the top 199 contestants achieving i points for the problem and n_{-1} is the number of those not submitting solutions.

A1. (156, 23, 4, 0, 0, 0, 0, 0, 0, 0, 16, 0)

Solution. Consider a plane cross-section through a vertex of the cube and the axis of the cone shown in Figure 1. Let s be the side of the cube. Segment DE is a

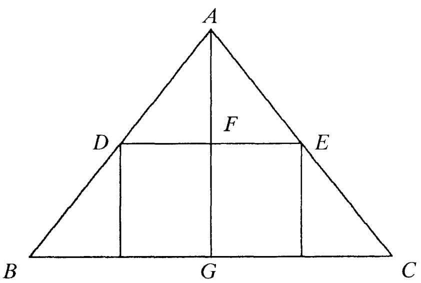


Figure 1.

diagonal of the top of the cube and has length $\sqrt{2}s$. By similarity of triangles ADE and ABC, we have $DE/BC = AF/AG$, or $\sqrt{2}s/2 = (3 - s)/3$. Hence

$$s = \frac{6}{3\sqrt{2} + 2} = \frac{3\sqrt{2}}{\sqrt{2} + 3} = \frac{9\sqrt{2} - 6}{7}.$$

A2. (103, 35, 26, 0, 0, 0, 0, 0, 9, 8, 12, 6)

Solution. Suppose the beginning and ending angles of s are α and β, respectively, where angles are measured counterclockwise from the positive horizontal axis.

$$A = \int_{\cos \beta}^{\cos \alpha} \sqrt{1 - x^2}\, dx = \int_{\beta}^{\alpha} \sqrt{1 - \cos^2 u}\,(-\sin u)\, du = \int_{\alpha}^{\beta} \sin^2 u\, du.$$

Similarly

$$B = \int_{\sin\alpha}^{\sin\beta} \sqrt{1-y^2}\, dy = \int_{\alpha}^{\beta} \cos^2 u\, du.$$

Then

$$A + B = \int_{\alpha}^{\beta} (\sin^2 u + \cos^2 u)\, du = \int_{\alpha}^{\beta} du = \beta - \alpha.$$

A3. (82, 34, 2, 0, 0, 0, 0, 0, 5, 0, 39, 37)

Solution. Assume $f(x)f'(x)f''(x)f'''(x) < 0$ for all x. By continuity, each of f, f', f'', f''' has constant sign. By replacing $f(x)$ by $f(-x)$, $-f(x)$, or $-f(-x)$, if necessary, we may assume that $f(x) > 0$ and $f'(x) > 0$ for all x. In either case, choosing $g = f$ or f' as necessary, we have a function g such that $g(x) > 0$, $g'(x) > 0$, and $g''(x) < 0$ for all x. Since g is strictly increasing and bounded below, there is a constant $C > 0$ such that $g(x) \to C^+$ as $x \to -\infty$. This horizontal asymptote forces g to be concave upwards at some point x, in contradiction to our assumption $g''(x) > 0$.

To be precise, fix x_1, and let $m = g'(x_1) > 0$. Since $\lim_{x_2 \to -\infty}(g(x_1) - g(x_2))/(x_1 - x_2) = 0$, there is an x_2, $x_2 < x_1$, such that $0 < (g(x_1) - g(x_2))/(x_1 - x_2) < m/2$. By the Mean Value Theorem, there is an x_3 strictly between x_2 and x_1 with $g'(x_3) = (g(x_1) - g(x_2))/(x_1 - x_2) < m/2$. Then there is an x_4 between x_3 and x_1 for which

$$g''(x_4) = \frac{g'(x_1) - g'(x_3)}{x_1 - x_3} > \frac{m - \frac{m}{2}}{x_1 - x_3} > 0.$$

A4. (39, 27, 52, 0, 0, 0, 0, 0, 49, 7, 14, 11)

Solution. A_n is divisible by 11 precisely when n is one more than a multiple of 6.

We know that a number is divisible by 11 if and only if the alternating sum of its digits is zero. So, let d_n be the alternating sum of the digits in A_n starting from the left. Then $d_n = d_{n-1} + (-1)^{r_{n-1}} d_{n-2}$, where r_n is the number of digits in A_n. Clearly, $r_n = F_n$ the nth Fibonacci number. The Fibonacci numbers have parity pattern $((-1)^{F_n})$: $-1, -1, 1, -1, -1, 1, \ldots$. Thus, the rules for calculating d_n are $d_1 = 0$, $d_2 = 1$, $d_3 = d_2 - d_1 = 1$, $d_4 = d_3 + d_2 = 2$, $d_5 = d_4 - d_3 = 1$, $d_6 = d_5 - d_4 = -1$, $d_7 = d_6 + d_5 = 0$, $d_8 = d_7 - d_6 = 1$, and so forth in this repeating pattern.

A5. (85, 24, 15, 0, 0, 0, 0, 0, 5, 2, 8, 60)

Solution. Let D_1 be a disc of greatest radius. Select D_2 to be a disc of greatest radius that is disjoint from D_1. Proceeding inductively, suppose that $D_1, \ldots, D_{j-1}$ have been chosen. Select D_j to be a disc of greatest radius that is disjoint from $D_1, \ldots, D_{j-1}$.

The process eventually stops, producing a pairwise disjoint collection of discs, $D_1, \ldots, D_n$. We claim that $\bigcup_{j=1}^{n} 3D_j \supseteq E$. To prove this assertion, it suffices to see that $\bigcup_{j=1}^{n} 3D_j \supseteq \bigcup_{D \in \mathscr{F}} D$.

Let $D \in \mathscr{F}$. If D is one of the selected discs D_j then the assertion is trivial. If D is not one of the selected discs, then let D_m be the first disc in the sequence that intersects D. Then, by the way we chose the D_j, the radius of D_m is at least as great as the radius of D. But the triangle inequality implies that $3D_m \supseteq D$.

A6. (12, 1, 3, 0, 0, 0, 0, 0, 3, 9, 92, 79)

Solution. The set $\mathbf{Z}^2$ is preserved by 90° rotation about any of its points. Therefore, there exists a point C' in $\mathbf{Z}^2$ with $|BC'| = |BA|$, BC' perpendicular to BA such that C and C' belong to the same half-plane with respect to the line through A and B. If $C \neq C'$, then $|CC'| = s \geq 1$. Let $r = |AB|$. By rotating and translating the coordinate system, we may assume that $B = (0, 0)$, $A = (r, 0)$, $C' = (0, r)$, and $C = (0, r) + s(\cos\theta, \sin\theta)$, where θ is the angle to the line $C'C$ measured from the positive horizontal axis. Thus

$$\begin{aligned}(|AB| + |BC|)^2 &= \left(r + \sqrt{s^2\cos^2\theta + (r + s\sin\theta)^2}\right)^2 \\ &= 2r^2 + s^2 + 2rs\sin\theta + 2r\sqrt{s^2\cos^2\theta + (r + s\sin\theta)^2} \\ &\geq 2r^2 + 1 + 2rs\sin\theta + 2r(r + s\sin\theta) \\ &= 8\frac{r(r + s\sin\theta)}{2} + 1 = 8[ABC] + 1.\end{aligned}$$

B1. (112, 30, 30, 0, 0, 0, 0, 0, 3, 3, 12, 9)

Solution 1. Let $z = x + 1/x$. Then

$$\begin{aligned}\frac{(x + 1/x)^6 - (x^6 + 1/x^6) - 2}{(x + 1/x)^3 + (x^3 + 1/x^3)} &= \frac{z^6 - z^2(z^2 - 3)}{z^3 + z(z^2 - 3)} \\ &= z^3 - z(z^2 - 3) = 3z = 3(x + 1/x).\end{aligned}$$

Observe that $x + 1/x - 2 = (\sqrt{x} - 1/\sqrt{x})^2 \geq 0$, with equality if and only if $x = 1$, so that the minimum of our function is 6.

Solution 2. By direct expansion followed by long-division of polynomials,

$$\begin{aligned}\frac{(x + 1/x)^6 - (x^6 + 1/x^6) - 2}{(x + 1/x)^3 + (x^3 + 1/x^3)} &= \frac{3}{x}\left(\frac{2x^8 + 5x^6 + 6x^4 + 5x^2 + 2}{2x^6 + 3x^4 + 3x^2 + 2}\right) \\ &= \frac{3}{x}(x^2 + 1) = 3(x + 1/x).\end{aligned}$$

By the arithmetic-geometric mean inequality, $x + 1/x \geq 2\sqrt{x(1/x)} = 2$, with equality if and only if $x = 1$. It follows that the minimum value is 6 and occurs when $x = 1$.

B2. (82, 10, 9, 0, 0, 0, 0, 0, 3, 11, 38, 46)

Solution. With two reflections, one about the x-axis and the other about $y = x$, the shortest perimeter for such a triangle is seen to equal the distance between the points $(a, -b)$ and (b, a). This distance is $\sqrt{2(a^2 + b^2)}$.

B3. (44, 4, 3, 0, 0, 0, 0, 0, 7, 2, 62, 77)

Solution. Observe that twice the desired surface area equals the surface area of the whole sphere minus 5 spherical caps subtended by a "chord" with central angle

$2\pi/5$. The surface area of a spherical cap of angle α can be computed in spherical coordinates by

$$\int_0^{\alpha/2}\int_0^{2\pi}\sin\phi\, d\theta\, d\phi = 2\pi(1-\cos(\alpha/2)).$$

Thus, the desired area is

$$\tfrac{1}{2}(4\pi - 5(2\pi(1-\cos(\pi/5)))) = 5\pi\cos(\pi/5) - 3\pi.$$

That is, $A = -3\pi$, $B = 5\pi$, $\alpha = \pi/2$, $\beta = \pi/5$.

B4. (42, 9, 22, 0, 0, 0, 0, 0, 21, 28, 24, 53)

Solution. It is necessary and sufficient that $mn/(m,n)^2$ is even, where (m,n) denotes the greatest common divisor. Indeed, if $m = ac$ and $n = bc$, then $\lfloor i/ac\rfloor + \lfloor i/bc\rfloor$ depends only on $\lfloor i/c\rfloor$, so

$$\begin{aligned}\sum_{i=0}^{abc^2-1}(-1)^{\left\lfloor\frac{i}{ac}\right\rfloor+\left\lfloor\frac{i}{bc}\right\rfloor} &= c\sum_{i=0}^{abc-1}(-1)^{\left\lfloor\frac{i}{a}\right\rfloor+\left\lfloor\frac{i}{b}\right\rfloor}\\ &= c\sum_{k=0}^{c-1}(-1)^{k(a+b)}\sum_{i=0}^{ab-1}(-1)^{\left\lfloor\frac{i}{a}\right\rfloor+\left\lfloor\frac{i}{b}\right\rfloor}.\end{aligned}$$

If ab is odd then $a+b$ is even, so

$$\sum_{k=0}^{c-1}(-1)^{k(a+b)} = c$$

and the sum of ab terms, each equal to ±1, is nonzero. Conversely, if $a+b$ is odd, then

$$\left\lfloor\frac{i}{a}\right\rfloor + \left\lfloor\frac{ab-1-i}{a}\right\rfloor + \left\lfloor\frac{i}{b}\right\rfloor + \left\lfloor\frac{ab-1-i}{b}\right\rfloor = b-1+a-1 \equiv 1 \pmod 2.$$

Thus,

$$\sum_{i=0}^{ab-1}(-1)^{\left\lfloor\frac{i}{a}\right\rfloor+\left\lfloor\frac{i}{b}\right\rfloor} = \sum_{i=0}^{ab/2-1}(-1)^{\left\lfloor\frac{i}{a}\right\rfloor+\left\lfloor\frac{i}{b}\right\rfloor} + (-1)^{\left\lfloor\frac{ab-1-i}{a}\right\rfloor+\left\lfloor\frac{ab-1-i}{b}\right\rfloor} = 0.$$

B5. (55, 2, 29, 0, 0, 0, 0, 0, 8, 0, 29, 76)

Solution. We have

$$\sqrt{N} = \sqrt{\frac{10^{1998}-1}{9}} = \frac{10^{999}}{3}(1-10^{-1998})^{1/2}.$$

Taylor's theorem with remainder implies that

$$(1-10^{-1998})^{1/2} = 1 - \frac{10^{-1998}}{2} + \epsilon, \quad \epsilon < \frac{10^{-3996}}{8}.$$

Thus, the first digit after the decimal, point in

$$10^{999}\sqrt{N} = \frac{10^{1998}-1}{3} + \frac{1}{6} + \frac{10^{1998}\epsilon}{3}$$

is 1.

B6. $(25, 8, 8, 0, 0, 0, 0, 0, 4, 5, 31, 118)$

Solution. Let $n = 4m^2$. Then

$$\begin{aligned} n^3 + an^2 + bn + c &= 64m^6 + 16am^4 + 4bm^2 + c \\ &= (8m^3 + am)^2 + (4b - a^2)m^2 + c. \end{aligned}$$

If $m \gg 0$, $|(4b - a^2)m^2 + c| < 8m^3 + am$, so $n^3 + an^2 + bn + c$ can be a square only if $(4b - a^2)m + c = 0$. This can happen for only one value of m unless $4b - a^2 = c = 0$, and in this case $n^3 + an^2 + bn + c$ is a square if and only if n is square.

Santa Clara University, Santa Clara, CA 95053-0290
Santa Clara University, Santa Clara, CA 95053-0290
St. Olaf College, Northfield, MN 55057-1098

UNDERSTANDING THE METRIC SYSTEM

1 million microphones = 1 megaphone
1 million bicycles = 2 megacycles
2000 mockingbirds = 2 kilomockingbirds
10 cards = 1 decacards
1/2 lavatory = 1 demijohn
1 millionth of a fish = 1 microfiche
453.6 graham crackers = 1 pound cake
10 rations = 1 decoration
10 millipedes = 1 centipede
3-1/3 tridents = 1 decadent
10 monologues = 5 dialogues
2 monograms = 1 diagram
8 nickels = 2 paradigms

Plucked from the Internet by Russ Hood, Rio Linda, CA

NOTES

Edited by **Jimmie D. Lawson and William Adkins**

On the Nelson Unit Distance Coloring Problem

Carsten Thomassen

In 1950 Nelson raised the problem of coloring the Euclidean plane in such a way that no two points of distance 1 receive the same color. How many colors are needed? This problem was often mentioned in Paul Erdős' famous lectures on unsolved combinatorial problems. The history of the problem is described in [**2**] and [**3**]. Clearly, three colors are needed. To see that four colors are needed, we consider seven points $x_1, x_2, \dots, x_7$ in the Euclidean plane such that the following pairs are of distance $1: x_1x_2, x_1x_3, x_1x_4, x_3x_4, x_2x_5, x_2x_6, x_5x_6, x_3x_7, x_4x_7, x_5x_7, x_6x_7$. It follows from the theorem of de Bruijn and Erdős [**1**] that the number of colors needed for the whole plane is the maximum number of colors needed for the finite subsets. The half-century old upper bound 7 is obtained by drawing an appropriate graph in the plane such that each face (region) is bounded by a cycle of (Euclidean) diameter less than one and then coloring each face and part of the boundary by the same color in such a way that only faces of distance > 1 receive the same color. We prove that colorings of this type always need at least 7 colors. More generally, 7 colors are needed for any surface and any metric of large diameter provided there are no short noncontractible curves and no short contractible curves whose interior have large area.

The upper bound 7 is obtained from a hexagonal tiling of the plane such that the hexagons are regular and of diameter slightly less than 1. All points inside a hexagon are colored with the same color. Two hexagons are colored differently if the distance between them is less than one. A coloring of this type will be called *nice*. More generally, we consider any metric space S, d such that S is a surface, i.e., S is an arcwise connected Hausdorff space such that each element of S has a neighborhood homeomorphic to an open disc in the Euclidean plane. We let G be a connected graph on S, i.e., the vertices of G are elements of S, and the edges of G are simple arcs on S that are pairwise disjoint except at a common vertex. Moreover, we assume that each face (i.e., arcwise connected component of $S \setminus G$) has diameter less than 1, is homeomorphic to a disc, and is bounded by a cycle in G. Now a *nice coloring* of S obtained from G is a coloring such that each color class is the union of faces (and part of their boundaries) such that the distance between any two of these faces is greater than 1. We define the *area* of subset A of S as the maximum number of pairwise disjoint open discs of radius $\frac{1}{2}$ that are contained in A. (If this maximum does not exist we say that A has infinite area.) We say that a simple closed curve C is *contractible* if $S \setminus C$ has precisely two arcwise connected components such that one of them is homeomorphic to an open disc in the Euclidean plane. That component is called the *interior* of C and is denoted *int*(C). (If S is a sphere, then *int*(C) denotes any component of $S \setminus C$ of

smallest area). We prove that every nice coloring of S needs at least 7 colors provided there exists a natural number k such that (i), (ii), (iii) below hold.

(i) Every noncontractible simple closed curve has diameter at least 2.
(ii) If C is a simple closed curve of diameter less than 2, then the area of $int(C)$ is at most k.
(iii) The diameter of S is at least $12k + 30$.

If any of these conditions (i), (ii), (iii) is dropped, then the number of colors needed may decrease. Thus a thin two-way infinite cylinder has a nice 6-coloring, which shows that (i) cannot be omitted. Similarly, a thin one-way infinite cylinder (with a small disc pasted on the boundary of the cylinder to form the bottom) shows that (ii) cannot be omitted. Finally, (iii) cannot be omitted since any sphere of diameter less than 1 has a nice coloring in two colors.

D. R. Woodall **[5]** (see also **[4]** for a correction) has obtained a 6-color theorem related to the 7-color theorem in our Theorem 1.

2. A lemma on degrees in graphs. A *graph* G is a set $V(G)$ of elements called *vertices* and a set $E(G)$ of unordered pairs xy of vertices called *edges*. If the edge xy is present we say that *xy joins* x and y and that x and y are *neighbors*. The number of neighbors of x is the *degree* of x. A *path* from x to y is a graph consisting of distinct vertices $x_1, x_2, \dots, x_n$ and the edges $x_1x_2, x_2x_3, \dots, x_{n-1}x_n$ where $x_1 = x$, $x_n = y$. If we add the edge x_nx_1 we obtain a *cycle*. If x is a vertex, then $D_1(x)$ is the set of neighbors of x. More generally, if $n \geq 2$, then $D_n(x)$ is the set of vertices in $V(G)\setminus[\{x\} \cup D_1(x) \cup \cdots \cup D_{n-1}(x)]$ having a neighbor in $D_{n-1}(x)$. The subgraph of G induced by $\{x\} \cup D_1(x) \cup D_2(x) \cup \cdots$ is the connected component of G containing x. The graph G is *connected* if G has only one connected component. G is *locally finite* if $D_1(x)$ is finite for each vertex x in G. G is *locally connected* if for each vertex x, the subgraph of G induced by $D_1(x)$ is connected. G is *locally Hamiltonian* if, for each x in $V(G)$, G has a cycle with vertex set $D_1(x)$. The graph of the icosahedron is locally Hamiltonian and has 12 vertices all of degree 5. No larger connected graph has these properties.

Lemma 1. *If G is a connected, locally finite, and locally Hamiltonian graph with at least* 13 *vertices, then G has a vertex of degree at least* 6.

Proof: If no vertex has degree at least 6 we pick a vertex x of maximum degree. Clearly x has degree at least 3. If x has degree 3, then the subgraph of G induced by $\{x\} \cup D_1(x)$ is the graph of the tetrahedron, because G has a cycle with vertex set $D_1(x)$. Since G has maximum degree 3, there is no vertex in $D_2(x)$. Since G is connected, G is the graph of the tetrahedron, contrary to the assumption that G has at least 13 vertices. If x has degree 4, then we consider a cycle in $D_1(x)$ and conclude that each vertex y in $D_1(x)$ has at most one neighbor z in $D_2(x)$. Since G has a cycle with vertex set $D_1(y)$, z has at least three neighbors in $D_1(x)$. So, there are at most 4 edges from $D_1(x)$ to $D_2(x)$, and every vertex in $D_2(x)$ has at least three neighbors in $D_1(x)$. Hence $D_2(x)$ has at most one vertex z. Since G has a cycle with vertex set $D_1(z)$, it follows that $D_3(x) = \emptyset$. So, G has at most 6 vertices, a contradiction. We may therefore assume that x has degree 5.

Each vertex y in $D_1(x)$ has at most two neighbors in $D_2(x)$, because a cycle with vertex set $D_1(y)$ shows that y has at least two neighbors in $D_1(x)$, Since G has a cycle with vertex set $D_1(y)$, every neighbor z of y in $D_2(x)$ has at least two neighbors in $D_1(x)$. Now z cannot have two or more neighbors in $D_3(x)$ because then a cycle with vertex set $D_1(z)$ shows that z has at least two neighbors in

$D_2(x)$, that is, z has a total of at least 6 neighbors, a contradiction. So z has at most one neighbor in $D_3(x)$ and that neighbor has at least three neighbors in $D_2(x)$. Since there are at most 10 edges from $D_1(x)$ to $D_2(x)$, and every vertex in $D_2(x)$ has at least two neighbors in $D_1(x)$, it follows that $D_2(x)$ has at most 5 vertices. Hence there are at most 5 edges from $D_2(x)$ to $D_3(x)$. Since each vertex in $D_3(x)$ has at least three neighbors in $D_2(x)$, it follows that $D_3(x)$ has at most one vertex, and $D_4(x) = \emptyset$. Hence G has at most 12 vertices, a contradiction that completes the proof. ∎

3. A 7-color theorem

Theorem 1. *Let G be a connected graph on a surface S satisfying (i), (ii), and (iii). Then every nice coloring needs at least 7 colors.*

Proof: Suppose (reductio ad absurdum) that there exists a coloring using at most 6 colors. We define the *map graph* $M = M(G, S)$ as the graph whose vertices are the faces of G such that two vertices in M are neighbors if and only if the corresponding facial cycles in G intersect. Consider any vertex x of M and let C_x be the corresponding facial cycle in G. We choose an orientation of C_x and let $x_1, x_2, \dots, x_k, x_1$ be the vertices in $D_1(x)$ listed in the order that they are encountered when we traverse C_x. We now explain the idea behind the proof. We consider first the particularly nice case where, for each vertex x, all vertices $x_1, x_2, \dots, x_k$ are distinct. In that case, M is locally Hamiltonian. Since the surface S is arcwise connected, it follows that M is connected. Since S has diameter greater than 13, M has more than 12 vertices, and hence, by Lemma 1, M has a vertex of degree at least 6. Now x and its neighbors must have distinct colors because x corresponds to a face of diameter < 1 on S. This contradiction completes the proof in the particularly nice case where M is locally Hamiltonian.

However, a vertex may appear several times in the sequence $x_1, x_2, \dots, x_k$ above, and some more careful analysis is needed. We omit those appearances (except possibly one) of x_i for which C_{x_i} and C_x have only a vertex in common. In other words, if x_i appears more than once in the new sequence, then we list only those appearances such that C_{x_i} and C_x share an edge. Then any two consecutive vertices in the sequence $x_1, x_2, \dots, x_k, x_1$ are neighbors in M and so M is locally connected. It follows that $M - x$ (that is, M with x and the edges incident with x removed) is connected. Moreover, if y is any other vertex of M, then $M - x - y$ is connected unless y appears twice in the sequence $x_1, x_2, \dots, x_k$, that is, C_x and C_y have at least two edges in common.

Consider now two vertices x and y such that C_x and C_y have at least two edges e and f in common (that is, $y = x_i = x_j$ for $1 \le i < j - 1 < k - 1$). Let R be a simple closed curve in the faces bounded by C_x and C_y such that R crosses each of e and f precisely once and has no other point in common with G. By (i), R is contractible. Hence $M - x - y$ is disconnected. We say that$\{x, y\}$ is a *2-separator* in M. For each vertex z in M such that C_z is in $int(R)$ and has color 1, we pick a point P_z in $int(C_z)$, By (ii), there are at most k points P_z and hence there are altogether at most $6k$ vertices z such that $int(C_z) \subseteq int(R)$. We define $int(M, x, y)$ as the subgraph of $M - x - y$ induced by all those vertices z in M such that C_z is in $int(R)$ for some R. Then each connected component of $int(M, x, y)$ has at most $6k$ vertices. Since S has diameter at least $12k + 3$ it follows that G has two vertices whose graph distance is at least $12k + 2$. Hence $M - x - y$ has some component that is not in $int(M, x, y)$. We claim that $M - x - y$ has precisely one such component, which we call $ext(M, x, y)$. To see this, let $e_1, e_2, \dots, e_m$ be the

edges in $C_x \cap C_y$ occuring in that cyclic order on C_x. Then $e_1, \ldots, e_m$ divide $D_1(x) \setminus \{y\}$ into m classes $A_1, A_2, \ldots, A_m$. By letting $\{e, f\} = \{e_i, e_{i+1}\}(1 \le i \le m)$ in the preceding argument, we conclude that for each $i = 1, 2, \ldots, m$, either $A_i \subseteq int(M, x, y)$ or $A_i \cap int(M, x, y) = \emptyset$. Since the former cannot hold for each $i \in \{1, 2, \ldots, m\}$, the latter must hold for some i, and hence the former holds for all other i in $\{1, 2, \ldots, m\}$. Summarizing, for any two vertices x, y in M, $M - x - y$ has precisely one connected component $ext(M, x, y)$ with more than $6k$ vertices.

If $\{u, v\}$ is a 2-separator in M such that either x or y or both is in $int(M, u, v)$, then clearly $int(M, x, y) \subset int(M, u, v)$. (To see this, we use the properties of M established previously and forget about S.) If no such 2-separator $\{u, v\}$ exists, then we say that $\{x, y\}$ is a *maximal 2-separator* and that *xy* is a *crucial edge*. Since each connected component of $int(M, x, y)$ has at most $6k$ vertices, then a maximal 2-separator exists (provided a 2-separator exists). Let H be the subgraph of M obtained by deleting $int(M, x, y)$ for each maximal 2-separator $\{x, y\}$. Then $H \neq \emptyset$. Moreover, H is connected since a shortest path in M between two vertices in H never uses vertices in $int(M, x, y)$. Similarly, H is locally connected. We claim that H is locally Hamiltonian. Consider again a vertex x in H and the sequence $x_1, x_2, \ldots, x_k, x_1$ in $D_1(x)$ (taken in M). If this sequence forms a Hamiltonian cycle in $D_1(x)$ in H, we have finished. By the definition of H, $k \ge 3$. So assume that $x_i = x_j$ where $1 \le i < j - 1 < k - 1$. Then $\{x, x_i\}$ is a 2-separator and the notation can be chosen such that $int(M, x, x_i\}$ contains all the vertices $x_{i+1}, x_{i+2}, \ldots, x_{j-1}$. We repeat this argument for each pair i, j such that $x_i = x_j$ where $1 \le i < j - 1 < k - 1$. Then the vertices in $x_1, x_2, \ldots, x_k, x_1$ that remain after we delete all vertices in the interiors of the 2-separators form a cyclic sequence with no repetitions. As H is connected and locally connected and has at least three vertices (by (iii)), the preceding reduced cyclic sequence has at least two distinct vertices. It cannot have precisely two vertices u, v because then $H - u - v$ is disconnected, and hence $M - u - v$ is disconnected (because M is obtained from H by "pasting graphs on edges of H"). Since one of the edges *xu* or *xv* is crucial (because $D_1(x)$ is smaller in H than in M), the maximality property of the 2-separator $\{x, u\}$ or $\{x, v\}$ implies that $ext(M, u, v)$ is the connected component of $M - u - v$ containing x. For each vertex z in that component, M has a path of length at most $6k$ from z to either x, u, or v. Hence M has diameter at most $12k + 1$, a contradiction that proves that H is locally Hamiltonian.

If H has a vertex x of degree at least 6 we have finished because x and its neighbors must have different colors in the nice coloring. So, we may assume that each vertex of H has degree at most 5. By Lemma 1, H has at most 12 vertices. Hence H has at most 30 edges. Since M is obtained from H by "pasting" $int(M, x, y\}$ on the crucial edge *xy* for each crucial edge of H, we conclude that the diameter of M is at most $12k + 29$, a contradiction to (iii).

REFERENCES

1. N. G. de Bruijn and P. Erdős, A colour problem for infinite graphs and a problem in the theory of relations. *Nederl. Akad. Wetensch. Proc. Ser. A* **54** (*Indag. Math.* **13** (1951) 371–373).
2. T. Jensen and B. Toft, *Graph Coloring Problems*, J. Wiley, New York, 1995.
3. A. Soifer, Chromatic number of the plane: a historical essay. *Geombinatorics* I (3) (1991) 13–15.
4. S. P. Townsend, Every 5-coloured map in the plane contains a monochrome unit, *J. Combinatorial Ser*, A **30** (1981) 114–115.
5. D. R. Woodall, Distances realized by sets covering the plane, *J. Combinatorial Theory Ser*. A **14** (1973) 187–200.

Technical University of Denmark, Building 303, DK-2800 Lyngby, Denmark
C.Thomassen@mat.dtu.dk

Descartes' Rule of Signs: Another Construction

David J. Grabiner

Descartes' Rule of Signs is a simple, classical bound on the number of positive roots of a polynomial, and analogously on the number of negative roots. In Descartes' own words, the rule is stated as follows: [**3**]

> An equation can have as many true [positive] roots as it has changes of sign, from $+$ to $-$ or from $-$ to $+$.

This statement is written in terms of sign changes of the coefficients, but the wording is very similar to the Intermediate Value Theorem, which says that a continuous function must have at least one root in an interval if it changes sign in that interval. This suggests a natural construction of the polynomial which achieves the bound by creating a correspondence between sign changes of its coefficients and sign changes of its value at designated points.

Other constructions that achieve the maximum number of roots are known [**2**]. In this MONTHLY, Anderson, Jackson, and Sitharam [**1**] give a different natural construction, which works as long as all signs are nonzero. They construct a polynomial by choosing the roots according to the desired signs; they then show that the coefficients of the polynomial have the correct signs. They also show that this polynomial can be modified to obtain a polynomial with the same sign sequence and any number of positive roots that is a positive even integer less than the maximum.

For our construction, we need only a single polynomial without signs attached to the coefficients; however we specify the signs, the value of the polynomial has the correct signs at the proper places.

Theorem 1. *Let $\sigma_0, \ldots, \sigma_n$ be any sequence of -1, 0, and $+1$. Then for any $k > n$, the polynomial*

$$p(x) = \sum_{j=0}^{n} \sigma_j k^{-j^2} x^j$$

has the maximum number of positive and negative roots allowed by Descartes' Rule of Signs.

Proof: If $\sigma_j \neq 0$, then at $x = k^{2i}$, the absolute value of the term of x^j is k^{2ij-j^2}. Thus, at $x = k^{2j}$, the absolute value of the term of x^{2j} is k^{j^2} and the absolute values of the other terms are all at most k^{j^2-1}. Since there are only n such terms and $k > n$, the term of x^{2j} is larger in absolute value than all the others combined, and thus the sign of $p(k^{2j})$ is the same as σ_j. Thus, if σ_i and σ_j are two consecutive opposite signs, the polynomial must have a root between $x = k^{2i}$ and $x = k^{2j}$.

Analogously, for negative roots, the term of x^{2j} is larger in absolute value than all the other terms at $x = -k^{2j}$, and thus if $(-1)^i\sigma_i$ and $(-1)^j\sigma_j$ are two

consecutive opposite signs in $p(-x)$, the polynomial must have a root between $x = -k^{2j}$ and $x = -k^{2i}$. ■

Since this theorem is valid for polynomials with zero coefficients, it is also easy to construct a polynomial with fewer roots than the maximum by making some of the coefficients very small and using the remaining signed coefficients to force the roots. In particular, we can make the number of positive roots short of the maximum by any even number.

Theorem 2. *Let $\sigma_0, \dots, \sigma_n$ be any sequence of -1, 0, and $+1$; we may assume $\sigma_0 \neq 0$ and $\sigma_n \neq 0$ since the nonzero roots are not affected by eliminating zeros at the ends of the sequence. Let $\tau_0, \dots, \tau_n$ be obtained by changing some of the internal σ_i to zero, but keeping τ_0 and τ_n nonzero. Then there is a polynomial $p(x)$ with sign sequence σ that has as many positive roots as there are sign changes in the τ sequence and as many negative roots as there are sign changes in the $(-1)^i\tau_i$ sequence.*

Corollary. *If the σ sequence has at least $2r$ sign changes, we can take the τ sequence by changing to zeros those signs that are opposite to σ_0 and precede the $2r$-th sign change. This gives a polynomial with $2r$ fewer positive roots than the number of sign changes in the σ sequence.*

Proof: Let $q(x)$ be the polynomial for the τ sequence; let its roots be $x_1, \dots, x_m$. The polynomial $q(x)$ has zero coefficients where certain signed coefficients are needed. We replace these zeros in $q(x)$ by sufficiently small terms $\sigma_i \delta x^i$ in $p(x)$ without affecting the number or signs of roots.

Since $q(x)$ is known to have its roots all in distinct intervals, it cannot have any double roots. Thus there is some ϵ_1 such that no root of $q'(x)$ is within ϵ_1 of a root of $q(x)$; we also require $\epsilon_1 < |x_i|$ for all i, which we can require since $q(0) = \tau_0 \neq 0$. Since $|q'(x)|$ is continuous, it is bounded on each interval $[x_i - \epsilon_1, x_i + \epsilon_1]$; let ϵ_2 be its minimum value on the union of all of these intervals. Finally, let ϵ_3 be the minimum of $\epsilon_1\epsilon_2$ and the value of $|q(x)|$ on the intervals $[-k^{2n}, x_1 - \epsilon_1]$, $[x_{i-1} + \epsilon_1, x_i - \epsilon_1]$, and $[x_m + \epsilon_1, k^{2n}]$.

Then if we add to $q(x)$ any differentiable function $f(x)$ that has $|f(x)| < \epsilon_3$ and $|f'(x)| < \epsilon_2$ on $[-k^{2n}, k^{2n}]$, then $f(x) + q(x)$ must still have one simple root in each interval $[x_i - \epsilon_1, x_i + \epsilon_1]$, since it changes sign in each such interval and its derivative does not change sign there. Also, $f(x) + q(x)$ cannot have any other root in $[-k^{2n}, k^{2n}]$, since $|f(x)| < \epsilon_3 < |q(x)|$ outside these intervals. Thus $f(x) + q(x)$ has the same number of roots as $q(x)$ in $[-k^{2n}, k^{2n}]$, with the same signs.

By the construction of $q(x)$ with leading term $\pm k^{-n^2}x^n$ and other terms whose absolute values are all at most $k^{-n^2-1}x^n$ for $x > k^{2n}$, we have $|q(x)| \geq (k - n)(k^{-n^2-1})x^n$ for all $x \geq k^{2n}$. Thus if $|f(x)| < (k - n)(k^{-n^2-1})x^n$ for all $x \geq k^{2n}$, then $f(x) + q(x)$ has no roots with absolute value greater than k^{2n}, and thus has the same number of positive and negative roots as q.

We can thus let $p(x) = f(x) + q(x)$, where

$$f(x) = \sum_{j=1}^{n-1} (\sigma_j - \tau_j)\delta x^j,$$

with δ sufficiently small to meet the conditions on f. This polynomial has the correct signs and the correct number of positive and negative roots. ■

Note that this technique does not allow us to obtain simultaneously all possible numbers of positive and negative roots. In fact, this turns out to be impossible; some combinations of positive and negative roots cannot be obtained at all. The

simplest impossible case is the sequence $+, 0, -, 0, +$, corresponding to the polynomial $ax^4 - bx^2 + c = 0$, which has $-x$ as a root if it has x as a root and thus cannot simultaneously have no positive and two negative roots.

Even if no signs are zero, it may not be possible to obtain simultaneously all admissible numbers of positive and negative roots. For example, the sequence $+, -, -, -, +$ has two positive and two negative sign changes. It is possible for a polynomial with this sign sequence to have two negative or zero positive real roots, but not both simultaneously. A fourth-degree polynomial with only two negative real roots for which the sum of the roots was positive could be factored as $a(x^2 + bx + c)(x^2 - sx + t)$ with $a, b, c, s, t > 0$, $s^2 < 4t$, and $b^2 \geq 4c$. The product of these factors is $a(x^4 + (b - s)x^3 + (t + c - bs)x^2 + (bt - cs)x + st)$. To get the correct sign sequence, we need $b < s$ and $bt < cs$, which gives $b^2 t < s^2 c$ and thus $b^2/c < s^2/t$. But we have $b^2/c \geq 4 > s^2/t$.

This counterexample provides a negative answer to the question raised in [**1**] whether it is possible to get a polynomial with an arbitrary sign sequence and any simultaneous numbers of positive and negative roots allowed by Descartes' Rule of Signs. This suggests a new conjecture: the only possible numbers of positive and negative roots are the maximum values permitted by Descartes' Rule of Signs in a sequence obtained by changing some of the internal signs to zeros as in Theorem 2. The above cases and the analogous $+, +, -, +, +$ confirm the conjecture for degree 4.

REFERENCES

1. Bruce Anderson, Jeffrey Jackson, and Meera Sitharam, Descartes' rule of signs revisited, *Amer. Math. Monthly* **105** (1998) 447–451.
2. I. Itenberg, and M. F. Roy, Multivariate Descartes' Rule, *Beiträge Algebra Geom.* **37** (1996) 337–346.
3. D. J. Struik, *ed. A source book in Mathematics 1200–1800*, Princeton University Press, 1986, pp. 89–93.

Bowling Green State University, Bowling Green, OH 43403
grabiner@math.bgsu.edu

Integrating Polynomials in Secant and Tangent

Jonathan P. McCammond

This note provides a relatively painless way to integrate arbitrary polynomials in secant and tangent without invoking integration by parts or anything beyond elementary polynomial and trigonometric identities. The techniques involved also introduce some of the ideas behind the construction of Laurent polynomials, although the manner in which they do so is rather indirect. We begin with a theorem that covers almost all possibilities.

Theorem 1. *For each polynomial $P(s, t)$ in two variables, there are polynomials F and G in one variable and a constant c such that*

$$\int P(\sec x, \tan x) \sec x \, dx = F(u) - G(v) + c \ln(u) + C$$

where $u = \sec x + \tan x$ and $v = \sec x - \tan x$.

Proof: Once we define $u = \sec x + \tan x$ and $v = \sec x - \tan x$, it is easy to check that $\sec x = (u+v)/2$, $\tan x = (u-v)/2$, and $uv = \sec^2 x - \tan^2 x = 1$. (The traditional construction of Laurent polynomials involves quotienting the two-variable polynomial ring by the two-sided principal ideal generated by the polynomial $xy - 1$. Since $uv = 1$, the polynomials in secant and tangent can be viewed as Laurent polynomials in the single variable $u = \sec x + \tan x$. This is an abstract algebra explanation for the simplifications that follow.) If we replace $\sec x$ and $\tan x$ with their equivalents in terms of u and v, then the polynomial P in $\sec x$ and $\tan x$ becomes a polynomial in u and v instead. Moreover, since $uv = 1$, we can replace any instance of uv by 1, quickly reducing any monomial containing both variables to one that contains at most a single variable. In other words, the resulting polynomial in u and v can always be written in the form $f(u)u + g(v)v + c$ where f and g are polynomials of a single variable and c is a constant.

Next consider the differentials du and dv. Since $du = \sec x \tan x + \sec^2 x\,dx$ and $dv = \sec x \tan x - \sec^2 x\,dx$, we find that $\sec x\,dx = (1/u)\,du = -(1/v)\,dv$. We are now ready to calculate the original integral.

$$\begin{aligned}\int P(\sec x, \tan x)\sec x\,dx &= \int (f(u)u + g(v)v + c)\sec x\,dx \\ &= \int f(u)\,du - \int g(v)\,dv + c\int \frac{1}{u}\,du \\ &= F(u) - G(v) + c\ln|u| + C\end{aligned}$$

where $F(u)$ and $G(v)$ represent the antiderivatives of the polynomials $f(u)$ and $g(v)$, respectively. ■

Example 2. Consider the integral $\int 16\sec^5 x\,dx$. We find that

$$\begin{aligned}P(s,t) &= 16s^4 \\ P\left(\frac{u+v}{2}, \frac{u-v}{2}\right) &= (u+v)^4 \\ &= u^4 + 4u^3v + 6u^2v^2 + 4uv^3 + v^4 \\ &= u^4 + 4u^2 + 6 + 4v^2 + v^4\end{aligned}$$

Thus $f(u) = u^3 + 4u$, $g(v) = 4v + v^3$, and $c = 6$; we immediately conclude that

$$\int 16\sec^5 x\,dx = \left(\frac{u^4}{4} + 2u^2\right) - \left(2v^2 + \frac{v^4}{4}\right) + 6\ln|u| + C$$

where $u = \sec x + \tan x$ and $v = \sec x - \tan x$. The standard approach would involve performing integration by parts twice in order to reduce the exponent of the integrand.

Theorem 1 is nearly comprehensive in the sense that the only monomials in secant and tangent that are not covered are the constant term and those of the form $\tan^n x$. Closed forms for these integrals exist [**1**], but a little bit of trigonometry brings them within the reach of Theorem 1. Consider, for example, the

monomial $\tan^7 x$. By repeatedly applying the identity $\tan^2 x = \sec^2 x - 1$ we see that

$$\begin{aligned}\tan^7 x &= \tan^5 x \sec^2 x - \tan^5 x \\ &= \tan^5 x \sec^2 x - \tan^3 x \sec^2 x + \tan^3 x \\ &= \tan^5 x \sec^2 x - \tan^3 x \sec^2 x + \tan x \sec^2 x - \tan x\end{aligned}$$

In the final expression, the first three terms can be integrated using Theorem 1, so that only $\tan x$ remains. More generally, given any polynomial in secant and tangent the pure powers of tangent can be modified in this way so that the result is the sum of a constant term, a constant multiple of $\tan x$, and a polynomial to which Theorem 1 can be applied. This completes the proof of the following result.

Theorem 3. *For each polynomial $P(s,t)$ in two variables, there are polynomials F and G in one variable and constants a, b, and c such that*

$$\int P(\sec x, \tan x)\, dx = F(u) - G(v) + a \ln|u| - b \ln|\cos x| + cx + C$$

where $u = \sec x + \tan x$ and $v = \sec x - \tan x$.

REFERENCE

1. W. Koepf, A. Ben-Israel, and R. P. Gilbert, *Mathematik mit Derive*, Vieweg, Braunschweig/Wiesbaden, 1993.

Texas A & M University, College Station, TX 77843
jon.mccammond@math.tamu.edu

THE EVOLUTION OF...

Edited by **Abe Shenitzer**

Mathematics and Statistics, York University, Toronto, Ontario M3J 1P3, Canada

Field Theory: From Equations to Axiomatization

Part II

Israel Kleiner

7. THE ABSTRACT DEFINITION OF A FIELD. The developments we have been describing thus far lasted close to a century. They gave rise to important "concrete" theories—Galois theory, algebraic number theory, algebraic geometry—in which the (at times implicit) field concept played a central role. At the end of the 19th century abstraction and axiomatics were "in the air." For example, Pasch (1882) gave axioms for projective geometry, stressing for the first time the importance of undefined notions, Cantor (1883) defined the real numbers essentially as equivalence classes of Cauchy sequences of rationals, and Peano (1889) gave his axioms for the natural numbers. In algebra, von Dyck (1882) gave an abstract definition of a group that encompassed both finite and infinite groups (about thirty years earlier Cayley had defined a *finite* group), and Peano (1888) gave a definition of a finite-dimensional vector space, though this was largely ignored by his contemporaries. The time was propitious for the abstract field concept to emerge. Emerge it did in 1893 in the hands of Weber (of Dedekind-Weber fame).

Weber's definition of a field appeared in his 1893 paper "Die allgemeinen Grundlagen der Galois'schen Gleichungstheorie" [**15**], in which he aimed to give an abstract formulation of Galois theory [**8**, p. 136]:

> In the following an attempt is made to present the Galois theory of algebraic equations in a way which will include equally well all cases in which this theory might be used. Thus we present it here as a direct consequence of the group concept illuminated by the field concept, as a formal structure completely without reference to any numerical interpretation of the elements used.

Weber's presentation of Galois theory is indeed very close to the way the subject is taught today. His definition of a field, preceded by that of a group, is as follows [**15**, pp. 526–527]:

> A group becomes a field if two types of composition are possible in it, the first of which may be called *addition*, the second *multiplication*. The general determination must be somewhat restricted, however.
> 1. We assume that both types of composition are commutative.
> 2. Addition shall generally satisfy the conditions which define a group.

3. Multiplication is such that

$a(-b) = -(ab)$

$a(b + c) = ab + ac$

$ab = ac$ implies $b = c$, unless $a = 0$

Given b and c, $ab = c$ determines a, unless $b = 0$.

Although the associative law under multiplication is missing, and the axioms are not independent, they are of course very much in the modern spirit. As examples of his newly defined concept Weber included the number fields and function fields of algebraic number theory and algebraic geometry, respectively, but also Galois's finite fields and Kronecker's "congruence fields" $K[x]/(p(x))$, K a field, $p(x)$ irreducible over K.

Weber proved (often reproved, after Dedekind) various theorems about fields, which later became useful in Artin's formulation of Galois theory, and which are today recognized as basic results of the theory. Among them are [**8**], [**10**]:

(i) Every finite algebraic extension of a field is simple (that is, it is generated by a single element).
(ii) Every polynomial over a field has a splitting field.
(iii) If $F \subseteq F(a) \subseteq F(b)$, then $(F(a): F)$ divides $(F(b): F)$, where for fields K and E with $E \subseteq K$, $(K : E)$ denotes the dimension of K as a vector space over E.

It should be emphasized that it was not Weber's aim to study fields as such, but rather to develop enough of field theory to give an abstract formulation of Galois theory [**11**]. In this he succeeded admirably. His paper, and somewhat later his two-volume *Lehrbuch der Algebra*, exerted considerable influence on the development of abstract algebra [**3**].

8. HENSEL'S *P*-ADIC NUMBERS. In an 1899 article entitled "New foundations of the theory of algebraic numbers," Hensel began a life-long study of p-adic numbers. Inspired by the work of Dedekind-Weber, Hensel took as his point of departure the analogy between function fields and number fields. Just as power series are useful for a study of the former, Hensel introduced p-adic numbers to aid in the study of the latter [**10**, II, p. 19]:

> The analogy between the results of the theory of algebraic functions of one variable and those of the theory of algebraic numbers suggested to me many years ago the idea of replacing the decomposition of algebraic numbers, with the help of ideal prime factors, by a more convenient procedure that fully corresponds to the expansion of an algebraic function in power series in the neighborhood of an arbitrary point.

Indeed, in the neighborhood of a given point α every algebraic function of a complex variable can be represented as an infinite series of integral and rational powers of $z - \alpha$, as Weierstrass had shown. The elements of Hensel's *field of p-adic numbers* are formal power series $\Sigma_n^\infty a_k p^k$, where $a_k \in Z_p$ and $n \in Z$. And just as every element of an algebraic function field can be identified with the set of its expansions at all points of the Riemann surface on which it is defined, so every element of an algebraic number field is identified with the set of its representations in the field of p-adic numbers $\Sigma_n^\infty a_k p^k$ for every prime p [**2**, p. 111].

In a 1907 book, Hensel introduced topological notions in his p-adic fields and applied the resulting p-adic analysis in algebraic number theory. The p-adic numbers proved extremely useful not only there but also in algebraic geometry [**4**], [**7**]. They were also influential in motivating the abstract study of rings and fields [**3**].

9. STEINITZ. The last major event in the evolution of field theory that we describe is Steinitz's great work of 1910 [**13**]. But first some background.

Algebra in the 19th century was by our standards concrete. It was connected in one way or another with the real or complex numbers. For example, some of the great contributors to 19th-century algebra, mathematicians whose ideas shaped the algebra of the 20th century, were Gauss, Galois, Jordan, Kronecker, Dedekind, and Hilbert, and their algebraic work dealt with quadratic forms, cyclotomy, permutation groups, ideals in rings of algebraic number fields and algebraic function fields, and invariant theory. All of these subjects were related in one way or another to the real or complex numbers.

At the turn of the 20th century the axiomatic method began to take hold as an important mathematical tool. Hilbert's *Foundations of Geometry* of 1899 was very influential in this respect (see also our Section 7). Noteworthy also is the American school of axiomatic analysis, as exemplified in the works of Dickson, Huntington, E. H. Moore, and Veblen. In the first decade of the 20th century these mathematicians began to examine various axiom systems for groups, fields, associative algebras, projective geometry, and the algebra of logic. Their principal aim was to study the independence, consistency, and completeness of the axioms defining any one of these systems. Also relevant were Hilbert's axiomatic characterization in 1900 of the field of real numbers and Huntington's like characterization in 1905 of the field of complex numbers [**1**], [**3**].

Steinitz's groundbreaking 150-page paper "Algebraische Theorie der Körper" of 1910 initiated the abstract study of fields as an independent subject [**13**]. While Weber *defined* fields abstractly, Steinitz *studied* them abstractly.

Steinitz's immediate source of inspiration was Hensel's p-adic numbers [**3**, p. 194]:

> I was led into this general research especially by Hensel's *Theory of Algebraic Numbers*, whose starting point is the field of p-adic numbers, a field which counts neither as a field of functions nor as a field of numbers in the usual sense of the word.

More generally, Steinitz's work arose out of a desire to delineate the abstract notions common to the various contemporary theories of fields: fields in algebraic number theory, in algebraic geometry, and in Galois theory, p-adic fields, and finite fields. His goal was a comprehensive study of *all* fields, starting from the field axioms [**3**, p. 195]:

> The aim of the present work is to advance an overview of all the possible types of fields and to establish the basic elements of their interrelations.

Quite a task! Steinitz's plan was to start from the simplest fields and to build up all fields from these. The basic concept that he identified to study the former is the *characteristic* of the field. Here are several of his fundamental results, nowadays

staples of field theory [**10**], [**13**]:

(i) Classification of fields into those of characteristic zero and those of characteristic p. The *prime fields*—the "simplest" fields—are Q and Z_p; one or the other is a subfield of every field.
(ii) Development of a theory of *transcendental extensions*, which became indispensable in algebraic geometry.
(iii) Recognition that it is precisely the *finite*, *normal*, *separable extensions* to which Galois theory applies.
(iv) Proof of the existence and uniqueness (up to isomorphism) of the *algebraic closure* of any field.

A description of all fields followed [**11**, p. 754]:

> Starting with an arbitrary prime field, by taking an arbitrary, purely transcendental extension followed by an arbitrary algebraic extension, we have a method of arriving at any field.

The notions of *transcendency base* and *degree of transcendence* of an extension field, both of which Steinitz introduced, played a crucial role here. Also important was the well-ordering principle (or, equivalently, the axiom of choice), whose use he acknowledged [**10**, II, p. 20]:

> Many mathematicians continue to reject the axiom of choice. The growing realization that there are questions in mathematics that cannot be decided without this principle is likely to result in the gradual disappearance of the resistance to it.

Steinitz's work was very influential in the development of abstract algebra in the 1920s and 1930s, as the following testimonials prove:

> Steinitz's paper was the basis for all [algebraic] investigations in the school of Emmy Noether (van der Waerden, [**14**, p. 162]).
>
> [Steinitz's work]... is not only a landmark in the development of algebra, but also... an excellent, in fact indispensable, introduction to a serious study of the new [modern] algebra (Baer and Hasse, [**13**, Preface]).
>
> Steinitz's work marks a methodological turning-point in algebra leading to ...'modern' or abstract algebra (Purkert and Wussing, [**11**, p. 754]).
>
> [Steinitz's work] can be considered as having given birth to the actual concept of Algebra (Bourbaki, [**2**, p. 83]).

10. A GLANCE AHEAD. We now list several major developments in field theory and related areas in the decades following Steinitz's fundamental work.

(a) *Valuation theory*. In 1913 Kürschak abstracted Hensel's ideas on p-adic fields by introducing the notion of a *valuation field*. He proved the existence of the completion of a field with respect to a valuation. In 1918 Ostrowski determined all valuations of the field Q of rational numbers. Valuation theory, which "forms a solid link between number theory, algebra, and analysis" [**7**, vol. II, p. 537], played fundamental roles in both algebraic number theory and algebraic geometry; see [**2**], [**4**], [**7**], [**14**].

(b) *Formally real fields*. In 1927 Artin and Schreier defined the notion of a *formally real field*, namely a field in which -1 is not a sum of squares. "One of [the] remarkable results [of the Artin-Schreier theory] is no doubt the discovery that the existence of an order relation on a field is linked to purely algebraic

properties of the field" [**2**, p. 92]: A field can be ordered if and only if it is formally real. The theory of formally real fields enabled Artin in the same year to solve *Hilbert's 17th Problem* on the resolution of positive definite rational functions into sums of squares [**7**, vol. II, p. 640].

(c) *Class field theory*. This is the study of finite extensions of an algebraic number field having an abelian Galois group. It is a beautiful synthesis of algebraic, number-theoretic, and analytic ideas, in which *Artin's Reciprocity Law* has a central place. Major strides were already made by Hilbert in his "Zahlbericht" (Report on Number Theory) of 1897. More modern aspects of the theory were developed by Artin, Chevalley, Hasse, Tagaki, and others; see [**5**].

(d) *Galois theory*. Artin set out his now-famous abstract formulation of Galois theory in lectures given in 1926 (but published only in 1938). In a 1950 talk he said [**8**, p. 144]:

> Since my mathematical youth I have been under the spell of the classical theory of Galois. This charm has forced me to return to it again and again, and try to find new ways to prove its fundamental theorems.

Extensions of the classical theory were given in various directions. For example, in 1927 Krull developed a Galois theory of *infinite field extensions*, establishing a one-one correspondence between subfields and "closed" subgroups, and thereby introducing topological notions into the theory. There is also a Galois theory for *inseparable field extensions*, in which the notion of derivation of a field plays a central role, and a Galois theory for *division rings*, developed independently by H. Cartan and Jacobson in the 1940s; see [**7**], [**16**].

(e) *Finite fields*. Finite field theory is a thriving subject of investigation in its own right, but it also has important uses in number theory, coding theory, geometry, and combinatorics; see [**6**], [**9**].

REFERENCES

1. G. Birkhoff, Current trends in algebra, *Amer. Math. Monthly* **80** (1973) 760–782, and corrections in **81** (1974) 746.
2. N. Bourbaki, *Elements of the History of Mathematics*, Springer-Verlag, 1984.
3. L. Corry, *Modern Algebra and the Rise of Mathematical Structures*, Birkhäuser, 1996.
4. D. Eisenbud, *Commutative Algebra with a View Toward Algebraic Geometry*, Springer-Verlag, 1995.
5. H. Hasse, History of class field theory, in *Algebraic Number Theory, Proceedings of an Instructional Conference*, ed. by J. Cassels and A. Fröhlich, Thompson Book Co., 1967, pp. 266–279.
6. K. Ireland and M. Rosen, *A Classical Introduction to Modern Number Theory*, 2nd ed., Springer-Verlag, 1982.
7. N. Jacobson, *Basic Algebra I, II*, W. H. Freeman, 1974 and 1980.
8. B. M. Kiernan, The development of Galois theory from Lagrange to Artin, *Arch. Hist. Exact Sci.* **8** (1971/72) 40–154.
9. R. Lidl and H. Niederreiter, *Introduction to Finite Fields and their Applications*, Cambridge University Press, 1986.
10. W. Purkert, Zur Genesis des abstrakten Körperbegriffs I, II, *Naturwiss., Techn. u. Med.* **8** (1971) 23–37 and **10** (1973) 8–20. Unpublished English translation by A. Shenitzer.
11. W. Purkert and H. Wussing, Abstract algebra, in *Companion Encyclopedia of the History and Philosophy of the Mathematical Sciences*, vol. I, ed. by I. Grattan-Guinness, Routledge, 1994, pp. 741–760.
12. J. H. Silverman and J. Tate, *Rational Points on Elliptic Curves*, Springer-Verlag, 1992.
13. E. Steinitz, *Algebraische Theorie der Körper*, 2nd ed., Chelsea, 1950.
14. B. L. van der Waerden, Die Algebra seit Galois, *Jahresber. Deutsch. Math. Verein.* **68** (1966) 155–165.
15. H. Weber, Die allgemeinen Grundlagen der Galois'schen Gleichungstheorie, *Math. Ann.* **43** (1893) 521–549.
16. D. Winter, *The Structure of Fields*, Springer-Verlag, 1974.

PROBLEMS AND SOLUTIONS

Edited by **Gerald A. Edgar, Daniel H. Ullman,** and **Douglas B. West**

with the collaboration of Paul T. Bateman, Mario Benedicty, Paul Bracken, Duane M. Broline, Ezra A. Brown, Richard T. Bumby, Glenn G. Chappell, Randall Dougherty, Roger B. Eggleton, Ira M. Gessel, Bart Goddard, Jerrold R. Griggs, Douglas A. Hensley, John R. Isbell, Robert Israel, Kiran S. Kedlaya, Murray S. Klamkin, Fred Kochman, Frederick W. Luttmann, Vania Mascioni, Frank B. Miles, Richard Pfiefer, Cecil C. Rousseau, Leonard Smiley, John Henry Steelman, Kenneth Stolarsky, Richard Stong, Charles Vanden Eynden, and William E. Watkins.

Proposed problems and solutions should be sent in duplicate to the MONTHLY *problems address on the inside front cover. Submitted problems should include solutions and relevant references. Submitted solutions should arrive at that address before April 30, 2000; Additional information, such as generalizations and references, is welcome. The problem number and the solver's name and address should appear on each solution. An acknowledgement will be sent only if a mailing label is provided. An asterisk (*) after the number of a problem or a part of a problem indicates that no solution is currently available.*

PROBLEMS

10760. *Proposed by Bruce Reznick, University of Illinois, Urbana, IL.* A function $f\colon \mathbb{N} \to \mathbb{C}$ is *completely multiplicative* if $f(1) = 1$ and $f(mn) = f(m)f(n)$ for all positive integers m and n. Find all completely multiplicative functions f with the property that the function $F(n) = \sum_{k=1}^{n} f(k)$ is also completely multiplicative.

10761. *Proposed by Fred Galvin, University of Kansas, Lawrence, KS.* Let G be a graph with n vertices. For each vertex v, let $f(v)$ be the maximum cardinality of an independent set of neighbors of v. Show that $\sum f(v) \le n^2/2$, where the sum is taken over all vertices of G.

10762. *Proposed by Leroy Quet, Denver, CO.* Let $x_1 = 1$, and for $m \ge 1$ let $x_{m+1} = (m + 3/2)^{-1} \sum_{k=1}^{m} x_k x_{m+1-k}$. Evaluate $\lim_{m\to\infty} x_m / x_{m+1}$.

10763. *Proposed by Jean Anglesio, Garches, France.* Let ABC be a triangle; let O be its circumcenter, H its orthocenter, I its incenter, N its Nagel point, and X, Y, Z its excenters. Let S be defined so that O is the midpoint of HS, and let T denote the midpoint of SN. It is known that the orthocenter and the nine-point center of triangle XYZ are I and O, respectively. Prove that
(**a**) the circumcenter of triangle XYZ is T; and
(**b**) the centroid of triangle XYZ is the centroid of SIN.

10764. *Proposed by Ray Redheffer, University of California, Los Angeles, CA.* Let $A = (a_{ij})$ be a real n-by-n matrix, and let x and y be real n-vectors satisfying $Ax = y$. Suppose that

$$\sum_{j\ne i} \max\{a_{ij}, 0\} < y_i \le a_{ii} + \sum_{j\ne i} \min\{a_{ij}, 0\}$$

for all $i \in \{1, 2, \ldots, n\}$. Show that $x_i > 0$ for all $i \in \{1, 2, \ldots, n\}$.

10765. *Proposed by Peter J. Ferraro, Roselle Park, NJ.* Let f_n be the nth Fibonacci number, defined by $f_1 = f_2 = 1$ and $f_{n+2} = f_{n+1} + f_n$ for $n \ge 1$. Fix positive integers k and n with $n \ge 2k + 1$. Prove that $\lfloor \sqrt[k]{f_n} \rfloor - \lfloor \sqrt[k]{f_{n-k}} + \sqrt[k]{f_{n-2k}} \rfloor$ is 0 unless f_n is a kth power, when it is 1.

10766. *Proposed by Szilárd András, Babeş-Bolyai University, Cluj-Napoca, Romania.* Let x, y, and z be nonnegative real numbers. Prove that

(a) $(x+y+z)^{x+y+z}x^x y^y z^z \le (x+y)^{x+y}(y+z)^{y+z}(z+x)^{z+x}$.

(b) $(x+y+z)^{(x+y+z)^2}x^{x^2}y^{y^2}z^{z^2} \ge (x+y)^{(x+y)^2}(y+z)^{(y+z)^2}(z+x)^{(z+x)^2}$.

SOLUTIONS

Cramer's Rule for Non-Square Matrices

10618 [1997, 768]. *Proposed by S. Lakshminarayanan, S. L. Shah, and K. Nandakumar, University of Alberta, Edmonton, Canada.* Let A be a real $m \times n$ matrix of full rank with $m < n$ and let b be a real $m \times 1$ matrix. For $1 \le i \le n$, define

$$x_i = \frac{\det(A_i^* A^T) - \det(A_i A_i^T)}{\det(AA^T)},$$

where A_i^* is obtained by replacing the ith column of A by b, and A_i is obtained by deleting the ith column of A. Show that $x = [x_1, \ldots, x_n]^T$ is a solution to the linear system $Ax = b$.

Solution by the GCHQ Problems Group, Cheltenham, U. K. We write $A^i\langle b\rangle$ instead of A_i^* to emphasize the role of the vector b; thus $A^i\langle 0\rangle$ indicates A with its ith column zeroed out. Observe that $A_i A_i^T = A^i\langle 0\rangle A^T$, by comparing corresponding entries.

Extend A to a nonsingular $n \times n$ matrix $\binom{A}{C}$, where C is an $(n-m) \times n$ matrix whose rows form an orthonormal basis for the orthogonal complement of the row space of A. That is, each row of C has norm 1 and is orthogonal to all other rows of $\binom{A}{C}$. We have

$$\begin{pmatrix} A \\ C \end{pmatrix}\begin{pmatrix} A \\ C \end{pmatrix}^T = \begin{pmatrix} AA^T & 0 \\ 0 & I \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} A^i\langle b\rangle \\ C \end{pmatrix}\begin{pmatrix} A \\ C \end{pmatrix}^T = \begin{pmatrix} A^i\langle b\rangle A^T & M \\ 0 & I \end{pmatrix},$$

where I is the $(n-m) \times (n-m)$ identity matrix and M is some $n \times (n-m)$ matrix. By substituting these computations into the definition of x_i, canceling the nonzero factor $\det \binom{A}{C}^T$, and using the linearity of the determinant in its ith column, we obtain

$$x_i = \frac{\det\left(\binom{A^i\langle b\rangle}{C}\binom{A}{C}^T\right) - \det\left(\binom{A^i\langle 0\rangle}{C}\binom{A}{C}^T\right)}{\det\left(\binom{A}{C}\binom{A}{C}^T\right)} = \frac{\det\binom{A^i\langle b\rangle}{C} - \det\binom{A^i\langle 0\rangle}{C}}{\det\binom{A}{C}} = \frac{\det\binom{A}{C}^i\!\left\langle\binom{b}{0}\right\rangle}{\det\binom{A}{C}},$$

By Cramer's rule, x is the solution to the linear system $\binom{A}{C}x = \binom{b}{0}$, and hence x is a solution to $Ax = b$.

Solved also by J. Fuelberth & A. Gunawardena, J. H. Lindsey II, M. Sharma & P. G. Poonacha (India), WMC Problems Group, and the proposers.

An Identity for Strongly Connected Digraphs

10620 [1997, 870]. *Proposed by James Propp, Massachusetts Institute of Technology, Cambridge, MA.* A *digraph* on a vertex set V is a subset $A \subseteq \{(v, w)\colon v, w \in V, v \ne w\}$ and is *strongly connected* if it is possible to get from any vertex a to every other vertex e by a finite succession of arcs $(a, b), (b, c), \ldots, (d, e)$ in A. For $n \ge 1$, let E_n (respectively, O_n) denote the number of strongly connected digraphs on the vertex set $V = \{1, 2, \ldots, n\}$ with an even (respectively odd) number of arcs. Show that $E_n - O_n = (n-1)!$ for all $n \ge 1$.

Solution I by the proposer, currently at University of Wisconsin, Madison, WI. The terminology of the problem statement is somewhat nonstandard. In common usage, a digraph is

a pair (V, A), in which A is the set of *arcs* in the digraph. Edges from a vertex to itself, and even multiple edges from one vertex to another, are ordinarily allowed; the digraphs considered here with no such edges are usually called *simple*. The "succession of arcs" described above is a *path* from a to e.

Let the *sign* of a digraph with arc set A be $(-1)^{|A|}$; the problem is to show that the sum s_n of signs of the strongly connected digraphs with vertex set $[n] = \{1, \ldots, n\}$ is $(n-1)!$. Let r_n be the sum of the signs of all the digraphs on the vertex set $[n]$ containing paths from n to all other vertices.

We prove first that $r_n = s_n - (n-1)s_{n-1}$ for $n > 1$. For $I \subseteq [n-1]$, let R_I be the set of digraphs on $[n]$ that contain paths from n to all other vertices and such that I is the set of vertices other than n that begin paths to vertex n. Let r_I be the sum of the signs of the digraphs in R_I, so $r_n = \sum r_I$.

When $|I| = n-1$, R_I is the set of strongly connected digraphs on $[n]$, and thus $r_I = s_n$.

When $|I| = n-2$, let j be the sole vertex of $[n-1] - I$. The signs of digraphs in R_I that have an arc from I to j sum to 0, since adding or deleting the arc (n, j) changes the sign without changing membership in this set. Thus we need to sum signs only for those digraphs in R_I that have no arc from I to j. Such digraphs consist of a strongly connected digraph on $[n] - \{j\}$ plus the arc (n, j), so the sum of their signs is $-s_{n-1}$. With $n-1$ choices for j, the sum of the contributions when $|I| = n-2$ is $-(n-1)s_{n-1}$.

When $|I| < n-2$, let i and j be the largest and second-largest elements of $[n-1] - I$. Adding or deleting the arc (i, j) changes the sign without changing membership in R_I. Hence these contributions cancel, and $r_I = 0$.

We have proved that $r_n = s_n - (n-1)s_{n-1}$. Since $s_1 = 1$, the proof is completed by proving that $r_n = 0$ for $n > 1$. For $I \subseteq [n-1]$, let Q_I denote the set of digraphs on $[n]$ in which I is the set of vertices other than n that are reachable by paths from vertex n. Let q_I be the sum of the signs of the digraphs in Q_I, so $r_n = q_{[n-1]}$.

When $|I| < n-1$, let i be the largest element of $[n-1] - I$. Adding or deleting the arc (i, n) changes the sign without changing membership in Q_I. Thus $q_I = 0$.

We conclude that $\sum_I q_I = r_n$. On the other hand, the sum of q_I over all I is the sum of the signs of all digraphs on $[n]$. This sum is 0, since adding or deleting the arc $(1, 2)$ changes the sign. Thus $r_n = 0$, as desired.

Composite solution II by Nikhil Bansal, Bombay, India, and the editors. Let $e(D)$ denote the number of arcs in a digraph D. The claim follows from comparing two expressions for $g(x)$, the exponential generating function for $a_n = \sum (-1)^{e(D)}$, where the sum is over *all* digraphs on $[n]$.

First, $g(x) = \sum_{n \geq 1} a_n x^n / n! = x$, since the number of digraphs with even size equals the number with odd size when $n > 1$, while $a_1 = (-1)^0 = 1$.

Alternatively, we group the contributions according to certain subdigraphs, counting those of even size minus those of odd size. The *strong components* of a digraph are the maximal strongly connected subdigraphs. The *condensation* D^* of a digraph D is the digraph obtained by contracting strong components to single points. The condensation of a digraph with k strong components is an acyclic digraph with k vertices.

To form a digraph on $[n]$ having k fixed strong components with vertex sets of sizes $n_1, \ldots, n_k$ respectively, we assign vertices to components, place a strongly connected digraph on each component, form an acyclic digraph on $[k]$ as the condensation, and expand edges of the condensation into sets of edges between the corresponding strong components. We divide by $k!$, because strong components are distinguished by the names of their vertices, but the names of the components are arbitrary.

The contributions to parity within each strong component and within the expansion of each edge of the acyclic digraph are independent; the final digraph has even size if and only if the number of odd contributions is even. Thus we take the product, over each set of edges

to be included, of the number of ways to include it with even size minus the number of ways to include it with odd size. Let $b_m = E_m - O_m$; this is the contribution for a strong component of order m.

At this point we have

$$\sum(-1)^{e(D)} = \sum_{k=1}^{n} \sum_{n_1+\cdots+n_k=n} \frac{n!}{n_1!\ldots n_k!}\frac{1}{k!}b_{n_1}\ldots b_{n_k}\sum_C \prod_{i,j\in\binom{[k]}{2}} (E_{ij} - O_{ij}),$$

where the inner sum is over acyclic digraphs C on $[k]$, the notation $\binom{[k]}{2}$ stands for the set of all 2-element subsets of $[k]$, and E_{ij} and O_{ij}, respectively, denote the number of ways to have an even or odd number of edges between components i and j in the expansion of C.

When i and j are not adjacent in C, there is no edge in the expansion, and $E_{ij} - O_{ij} = 1$. When i and j are adjacent, there must be at least one edge in the expansion, and all such edges agree in direction with the edge in C. Eliminating the empty set yields $E_{ij} - O_{ij} = -1$. With the factor -1 for each edge of C, the product is $(-1)^{e(C)}$.

To further simplify the formula, we claim that $\sum_C(-1)^{e(C)} = (-1)^{k-1}$. We define an involution on the acyclic digraphs that pairs up digraphs with sizes differing by 1, and we show that the only unpaired digraph is the digraph C_k with arc set $\{(k, j)\colon 1 \le j \le k-1\}$. A *source* is a vertex with no incoming arc; a *predecessor* of j is a vertex i such that (i, j) is an arc.

Every acyclic digraph C has at least one source. Let i be the least source vertex. When $i \neq k$, we add or delete the arc (i, k); it remains true that i is the least source. In the remaining digraphs, the only source vertex is k. For these, let j be the highest vertex having a predecessor other than k. Add or delete the arc (k, j). It remains true that k is the only source and that j is the highest vertex with a predecessor other than k. The only digraph in which j does not exist is C_k (with $k-1$ edges), which completes the claim.

The coefficient of $x^n/n!$ in our generating function is now

$$\sum_{k=1}^{n} \sum_{n_1+\cdots+n_k=n} \binom{n}{\{n_i\}} \prod b_{n_i} \frac{(-1)^{k-1}}{k!}.$$

Given the exponential generating function $f(x) = b_n x^n/n!$, this yields $g(x) = 1 - e^{-f(x)}$. Since $g(x) = x$, we obtain $f(x) = -\ln(1-x)$, and hence $b_k = (k-1)!$.

Editorial comment. The proposer's proof is adapted from an analysis of the unsigned case due to V. A. Liskovec (On a recurrence method of counting graphs with labelled vertices, *Soviet Math. Dokl.* 10 (1969) 242–256) and explicated by E. M. Wright (The number of strong digraphs, *Bull. London Math. Soc.* 3 (1971) 348–350). Problem 6673 [1991, 965; 1994, 686] in this MONTHLY is the analogous problem for undirected graphs. The proposer notes the following consequence. If the sign of a disjoint union of strongly connected (simple) digraphs is $(-1)^{e+k}$, where e is the total number of arcs and k is the number of components, then the sum of the signs of all n-vertex disjoint unions of strongly connected digraphs is 0. He asks whether there is a simple direct proof of this corollary. One might also ask for a simple signed involution to prove the original claim directly.

Solved also by R. J. Chapman (U. K.).

Simultaneous Squares from Arithmetic Progressions

10622 [1997, 870]. *Proposed by M. N. Deshpande, Nagpur, India.* Find infinitely many triples (a, b, c) of positive integers such that a, b, c are in arithmetic progression and such that $ab+1$, $bc+1$, and $ca+1$ are perfect squares.

Solution I by Hansruedi Widmer, Nussbaumen, Switzerland. Let $a_0 = 1$, $a_1 = 4$, and $a_{n+2} = 4a_{n+1} - a_n$ for $n \ge 0$, and set $b_n = 2a_{n+1}$ and $c_n = a_{n+2}$. We claim that

(a_n, b_n, c_n) meets the conditions of the problem. By the recurrence, $c_n - b_n = b_n - a_n$, and we have arithmetic progressions.

To prove that $a_n c_n + 1$ is a perfect square, we prove by induction on n that $a_n a_{n+2} + 1 = a_{n+1}^2$. For $n = 0$, we have $1 \cdot 15 + 1 = 4^2$. For $n > 0$, we use the recurrence and the induction hypothesis to compute

$$a_n a_{n+2} + 1 = 4a_n a_{n+1} - a_n^2 + 1 = 4a_n a_{n+1} - a_{n-1}a_{n+1} = a_{n+1}^2.$$

Next, we use this formula and the recurrence to compute

$$\begin{aligned} a_n b_n + 1 &= 2a_n a_{n+1} + 1 = 2a_n a_{n+1} + a_{n+1}^2 - a_n a_{n+2} = a_{n+1}^2 - a_n(a_{n+2} - 2a_{n+1}) \\ &= a_{n+1}^2 - a_n(2a_{n+1} - a_n) = (a_{n+1} - a_n)^2. \end{aligned}$$

Similarly $b_n c_n + 1 = (a_{n+2} - a_{n+1})^2$, so all desired properties hold.

Solution II by Zachary Franco, Butler University, Indianapolis, IN. In $\mathbb{Z}[\sqrt{3}]$, the norm $\|r + s\sqrt{3}\| = r^2 - 3s^2$ is multiplicative and satisfies $\|2 + \sqrt{3}\| = 1$. Therefore, $\|(2 + \sqrt{3})^n\| = 1$. For the expansion $(2+\sqrt{3})^n = r + s\sqrt{3}$, we thus have

$$3s^2 = r^2 - 1. \qquad (*)$$

For $n > 1$, the triple $(a, b, c) = (2s - r, 2s, 2s + r)$ is in arithmetic progression and satisfies $(2s - r)2s + 1 = (r - s)^2$, $(2s - r)(2s + r) + 1 = s^2$, and $2s(2s + r) + 1 = (r + s)^2$.

Editorial comment. The two solutions generate the same family of triples. Jan Kristian Haugland, John P. Robertson, and Ivan Vidav independently proved that this family contains all triples with $a \le b \le c$ that satisfy the conditions of the problem. Betsy Carper, Gary Hull, Lenny Jones, and Bonnie Wachhaus observed that for each triple in this family there is no fourth integer d such that a, b, c, d are in arithmetic progression and both $bd + 1$ and $cd + 1$ are perfect squares. Hence there is no quadruple of positive integers in arithmetic progression such that $ij + 1$ is a perfect square for all i, j in the quadruple.

David M. Bloom observed that this problem is related to problem 10238 [1992, 674; 1995, 275]. The published solution for that problem used the Pell equation $(*)$.

Solved also by I. Adler, M. Aissen, J. Anglesio (France), B. D. Beasley, J. C. Binz (Switzerland), D. M. Bloom, J. Bowring, J. T. Bruening, S. Byrd, R. J. Chapman (U. K.), J. Christopher, C. R. Diminnie, R. DiSario, H. Y. Far, J. K. Haugland (Norway), R. Heller, L. Jones et al., N. Komanda, J. Lee, N. F. Lindquist, C. Mack, L. G. Mans, A. S. Mittal, G. R. Mott, Y. Pan, M. Reekie, J. P. Robertson, H. Sedinger, N. C. Singer, W. R. Smythe, P. Trajovský (Czech Republic), I. Vidav (Slovenia), M. Vowe (Switzerland), X. Wang, C. H. Webster, P. Yiu, Anchorage Math Solutions Group, GCHQ Problems Group (U. K.), NSA Problems Group, NCCU Problems Group, SAS Maths Club (India), WMC Problems Group, and the proposer.

The Divisibility Poset Inside Itself

10623 [1997, 870]. *Proposed by Roy Barbara, Lebanese University, Fanar, Lebanon.* Let $P = \{1, 2, 3, \ldots\}$, and let $|$ be the usual divisibility relation on P. For any $S \subseteq P$ and $n \in P$, let $S + n = \{s + n : s \in S\}$.

(**a**) Can one construct a subset S of P such that the poset $(S, |)$ is isomorphic to $(P, |)$, $(S + 1, |)$ is isomorphic to $(P, \le)$, and $(S + 2, |)$ is isomorphic to $(P, |)$?

(**b**) For which integers $n \ge 1$ can one find a subset T of P such that $(T, |)$, $(T + n, |)$, and $(P, |)$ are isomorphic posets?

Solution to (**a**) *by Nasha Komanda, Central Michigan University, Mt. Pleasant, MI.* The answer is yes. First we prove for positive integers x, y and for integers $a \ne \pm 1$ that

$$\gcd(a^x - 1, a^y - 1) = |a^{\gcd(x,y)} - 1|. \qquad (*)$$

We may assume that $x \ge y$. From $a^x - 1 = a^{x-y}(a^y - 1) + (a^{x-y} - 1)$ we obtain $\gcd(a^x - 1, a^y - 1) = \gcd(a^y - 1, a^{x-y} - 1)$. Also $\gcd(x, y) = \gcd(y, x - y)$. Thus $(*)$ follows by induction on $x + y$. If also x and y are odd, then

$$\gcd(a^x + 1, a^y + 1) = \gcd((-a)^x - 1, (-a)^y - 1) = |(-a)^{\gcd(x,y)} - 1| = |a^{\gcd(x,y)} + 1|.$$

For positive integers a, x, y with $a > 1$, it follows that $(a^x - 1)|(a^y - 1)$ if and only if $x|y$. When x and y are odd, we also have $(a^x + 1)|(a^y + 1)$ if and only if $x|y$.

Now let $S = \{2^{2^x-1}\colon x \in P\}$. For $x, y \in P$,

$$(2^{2^x-1} - 1)|(2^{2^y-1} - 1) \iff (2^x - 1)|(2^y - 1) \iff x|y.$$

Thus $(S, |)$ is isomorphic to $(P, |)$. Also

$$(2^{2^x-1})|(2^{2^y-1}) \iff (2^x - 1) \leq (2^y - 1) \iff x \leq y.$$

Thus $(S + 1, |)$ is isomorphic to $(P, \leq)$. Since $2^x - 1$ and $2^y - 1$ are odd, we have

$$(2^{2^x-1} + 1)|(2^{2^y-1} + 1) \iff (2^x - 1)|(2^y - 1) \iff x|y.$$

Thus $(S + 2, |)$ is isomorphic to $(P, |)$.

Solution to (**b**) *by Robin J. Chapman, University of Exeter, Exeter, U. K.* Such a T exists for all $n \in P$. Let $O = \{1, 3, 5, \ldots\}$ be the set of odd positive integers. Let p_j denote the j-th smallest prime. Then $\phi : \prod_j p_j^{r_j} \to \prod_j p_{j+1}^{r_j}$ is an isomorphism from $(P, |)$ to $(O, |)$.

Given $n \in P$, let $T = nO = \{nm : m \in O\}$. Then $T + n = 2nP = \{2nk : k \in P\}$. The posets $(T, |)$ and $(O, |)$ are isomorphic, and $(T + n, |)$ and $(P, |)$ are isomorphic. We obtain $(T, |)$ isomorphic to $(T + n, |)$ by transitivity.

Part (**a**) solved also by R. J. Chapman (U. K.). Both parts solved also by J. Dawson (Australia), GCHQ Problems Group (U. K.), and the proposer.

An Equation Involving the Totient

10626 [1997, 871]. *Proposed by Florian Luca, Syracuse University, Syracuse, NY.* For a positive integer k, the number of positive integers less than k that are relatively prime to k is denoted $\phi(k)$.

(**a**) Show that if m and n are relatively prime positive integers, then $\phi(5^m - 1) \neq 5^n - 1$.

(**b**)* Find all positive integers m, n such that $\phi(5^m - 1) = 5^n - 1$.

Solution to (**a**) *by Nasha Komanda, Central Michigan University, Mt. Pleasant, MI.* We use the lemma proved in Problem **10623**: $\gcd(a^m - 1, a^n - 1) = a^{\gcd(m,n)} - 1$ when a, m, n are positive integers.

When $\gcd(m, n) = 1$, the lemma yields $\gcd(5^m - 1, 5^n - 1) = 4$. When $5^m - 1$ has the prime factorization $2^{e_0} \prod_{i=1}^{s} p_i^{e_i}$ with all exponents positive, we have $\phi(5^m - 1) = 2^{e_0-1} \prod_{i=1}^{s} p_i^{e_i-1}(p_i - 1)$. Suppose that $\phi(5^m - 1) = 5^n - 1$. Since $\gcd(5^m - 1, 5^n - 1) = 4$, we have $e_i = 1$ for $1 \leq i \leq s$.

If m is even, then $5^m - 1 = 25^{m/2} - 1$ is divisible by 8. Therefore, $e_0 \geq 3$. Since m and n are relatively prime, n is odd, so $5^n - 1 \equiv 4 \pmod 8$, which implies that $e_0 - 1 + s \leq 2$. Thus $e_0 = 3$ and $s = 0$, which yields the impossibility $5^m - 1 = 8$.

Thus m is odd, which yields $e_0 = 2$. Since $e_i = 1$ for $1 \leq i \leq s$, we have $5^m - 1 = 4 \prod_{i=1}^{s} p_i$. Thus $5^m \equiv 1 \pmod{p_i}$ for $1 \leq i \leq s$. Since m is odd, 5 is a quadratic residue modulo p_i. By the Quadratic Reciprocity Law, each p_i is a quadratic residue modulo 5. Thus $p_i \equiv \pm 1 \pmod 5$. If $p_i \equiv 1 \pmod 5$, then $p_i - 1 \equiv 0 \pmod 5$ and $5^n - 1 \equiv 0 \pmod 5$, which is impossible. Therefore, $p_i \equiv -1 \pmod 5$ for $1 \leq i \leq s$. Our formula for $5^m - 1$ now yields $-1 \equiv (-1)^{s+1} \pmod 5$, and hence s is even.

On the other hand, $e_i = 1$ for $1 \leq i \leq s$ also yields $5^n - 1 = 2 \prod_{i=1}^{s}(p_i - 1)$. With $p_i \equiv -1 \pmod 5$, this yields $-1 \equiv 2(-2)^s \pmod 5$. This requires $s \equiv 3 \pmod 4$, which contradicts our conclusion that s is even. Thus m can be neither odd nor even, and no solution exists.

Editorial comment. No solution was received to part (**b**). Roy Barbara provided partial results about the general equation $\phi(a^m - 1) = a^n - 1$, with additional partial results in

the special case $a = 5$. When a is even, there is no solution except $(a, m, n) = (2, 1, 1)$. When a is odd and at least 3, there is no solution when n is odd, when m is a power of 2, or when m is divisible by at least as high a power of 2 as n.

When $a = 5$, further exclusions restrict the possibilities for (m, n) to $\{(2^k p, 2^k q) : p$ is odd, q is even, and $0 \leq k \leq 4\}$. Furthermore, such a solution requires the five equations $\phi(5^{2^j p} - 1) = 5^{2^j q} - 1$ for $0 \leq j \leq 4$. This is highly restrictive; there may be no solution.

Part (**a**) solved also by H. Salle, GCHQ Problems Group, and the proposer.

Three Congruent Circles Between Two Triangles

10659 [1998, 366]. *Proposed by Jiro Fukuta, Shinsei-cho, Gifu-ken, Japan.* Let D, E, F be points in the interior of sides BC, CA, AB, respectively, of triangle $\triangle ABC$ such that the incircles of $\triangle AEF$, $\triangle BFD$, and $\triangle CDE$ are congruent, each having radius r. Let ρ, s, and K be the inradius, semiperimeter, and area of $\triangle ABC$, and ρ', s', and K' be the corresponding quantities for $\triangle DEF$.

(**a**) Prove that $\rho' = \rho - r$, $s' = (1 - r/\rho)s$, and $K' = (1 - r/\rho)^2 K$.

(**b**) Prove that, if $r = \rho/2$, then D, E, and F are midpoints of the sides of $\triangle ABC$.

Solution of part (**a**) *by GCHQ Problems Group, Cheltenham, U. K.* Let the incentres of $\triangle AEF$, $\triangle BFD$, and $\triangle CDE$ be O_1, O_2, and O_3 respectively. We first show that the triangles $\triangle O_1O_2O_3$ and $\triangle DEF$ have the same area and same perimeter.

Let the incircle of $\triangle AEF$ touch AF at T and EF at U. Let the incircle of $\triangle BDF$ touch BF at V and DF at W. Then $FT = FU$ and $FV = FW$, so $FU + FW = FT + FV = TV = O_1O_2$. Continuing this process around $\triangle DEF$ shows that it has the same perimeter as $\triangle O_1O_2O_3$.

Let G be the foot of the perpendicular from F to O_1O_2 and let H be the intersection of O_1O_2 and EF. Then Area$(\triangle FGO_1)$ = Area$(\triangle FTO_1)$ = Area$(\triangle FUO_1)$, so Area$(\triangle FGH)$ = Area$(\triangle O_1UH)$. Continuing in this manner shows that $\triangle DEF$ has the same area as $\triangle O_1O_2O_3$. The result is true whether or not line segments FG and O_2W intersect. Since the two triangles have the same area and the same perimeter, they must also have the same inradius, namely ρ'.

Now $\triangle O_1O_2O_3$ is homothetic to $\triangle ABC$, with the incentre of $\triangle ABC$ as the centre of homothety. It follows that $\rho' = \rho - r$. The scaling factor from $\triangle ABC$ to $\triangle O_1O_2O_3$ is ρ'/ρ, so $s' = (\rho'/\rho)s = (1 - r/\rho)s$ and $K' = (\rho'/\rho)^2 K = (1 - r/\rho)^2 K$ as required.

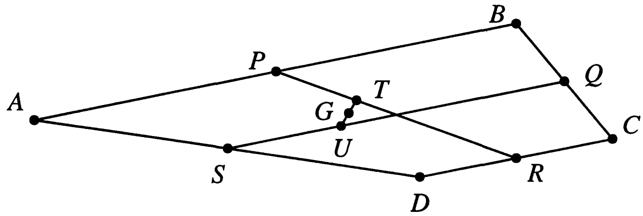
Solution of part (**b**) *by Călin Popescu, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.* For brevity, let us call a triple of points D, E, F with the properties described in the statement of the problem *r-adequate.* The midpoints D_0, E_0, F_0 of the sides BC, CA, AB, respectively, form a $\rho/2$-adequate triple; we have to show that it is the only such triple. To this end, let D, E, and F form a $\rho/2$-adequate triple, and note that $K'/K = 1/4$, by part (**a**). Observe that BD/BC, CE/CA, and AF/AB can be written as $1/2 + x$, $1/2 + y$, and $1/2 + z$, respectively, with x, y, and z either all in $(-1/2, 0]$ or all in $[0, 1/2)$. Indeed, if say $x < 0 < y$, then clearly the inradius of $\triangle CDE$ would be larger than the inradius of $\triangle CD_0E_0$, which is impossible since both are $\rho/2$. Writing K_A, K_B, K_C for the areas of $\triangle AEF$, $\triangle BFD$, $\triangle CDE$, respectively, we obtain $K_A/K = (AE/AC)(AF/AB) = (1/2 - y)(1/2 + z) = 1/4 + (z - y)/2 - yz$ and similarly $K_B/K = 1/4 + (x - z)/2 - zx$ and $K_C/K = 1/4 + (y - x)/2 - xy$. It follows that $K'/K = 1 - (K_A/K + K_B/K + K_C/K) = 1/4 + xy + yz + zx$. Since x, y, and z are either all nonnegative or all nonpositive, $K'/K = 1/4$ forces $x = y = z = 0$. Hence D, E, and F are indeed the midpoints of the sides of the triangle ABC.

Solved also by T. Hermann, N. Lakshmanan, G. Peng, C. R. Pranesachar (India), A. Sasane (The Netherlands), and the proposer. Part (**a**) solved also by R. Barbara (France), F. Bellot Rosado (Spain), J. Dou (Spain), P. E. Nuesch (Switzerland), C. Popescu (Belgium), W. Reyes (Chile), R. A. Simon (Chile), and I. Sofair.

Quadrilateral Center of Gravity

10662 [1998, 464]. *Proposed by Joseph D. E. Konhauser and Stan Wagon, Macalester College, St. Paul, MN.* Find a construction for the center of gravity of the edges of a quadrilateral.

Solution by the Con Amore Problems Group, Royal Danish School of Educational Studies, Copenhagen, Denmark. If G is the center of gravity of the edges of the quadrilateral $ABCD$ then G is also the center of gravity of particles with masses proportional to the lengths of the edges AB, BC, CD, DA placed at the midpoints P, Q, R, S of these edges. Construct these midpoints. The center of gravity for the particles at P and R is the point T on PR such that $PT : TR = CD : AB$, and the center of gravity for the particles at Q and S is the point U on QS with $QU : US = DA : BC$, so we construct the points T and U. If T and U coincide, we have $G = T = U$. If not, then G is the center of gravity of particles with masses proportional to the lengths of the line segments $AB + CD$ and $BC + DA$ placed at T and U, respectively. The sum of two line segments is constructed by placing them end to end. Thus G is the point on TU with $TG : GU = (BC + DA) : (AB + CD)$, and we construct this point.



Solved also by M. Benedicty, M. Boase (U. K.), G. D. Chakerian, R. J. Chapman (U. K.), S. S. Kim (Korea), J. H. Lindsey II, A. Nijenhuis, V. Pambuccian, C. R. Pranesachar (India), A. Sasane (The Netherlands), J. Schaer (Canada), Anchorage Math Solutions Group, GCHQ Problems Group, and the proposers.

Logarithmic Convexity of Stirling's Ratio

10680 [1998, 666]. *Proposed by Harold G. Diamond, University of Illinois, Urbana, IL.* For $x > 0$ set $g(x) = x \log\left(\Gamma(x+1)/(x^x e^{-x}\sqrt{2\pi x})\right)$. Show that g is concave down on $(0, \infty)$.

Solution by Nathaniel Grossman, University of California, Los Angeles, CA. It is enough to show that $g''(x) < 0$ when $x > 0$. We begin with Binet's second expression for the gamma function, which we write in the form

$$\log \Gamma(x+1) = (x + \frac{1}{2}) \log x - x + \frac{1}{2} \log(2\pi) + 2k(x), \qquad (*)$$

where $k(x) = \int_0^\infty \arctan(t/x)(e^{2\pi t} - 1)^{-1}\, dt$ (M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions,* Dover, New York, 1972, p. 258 (**6.1.50**)). From $(*)$ we find that $g(x) = 2xk(x)$, hence $g''(x) = 2\left(xk''(x) + 2k'(x)\right)$. Easily justified differentiation under the integral sign leads to

$$xk''(x) + 2k'(x) = -\int_0^\infty \frac{2t^3}{(x^2+t^2)^2(e^{2\pi t} - 1)}\, dt,$$

in which the right hand side is clearly negative.

Solved also by J. Anglesio (France), P. Bracken (Canada), D. Bradley, E. Camouzis (Greece), R. J. Chapman (U. K.), R. A. Groeneveld, D. Krug, O. P. Lossers (The Netherlands), R. Martin (U. K.), A. McD. Mercer (Canada), P. Simeonov, A. Stadler (Switzerland), and NCCU Problems Group.

REVIEWS

Edited by **Harold P. Boas**
Mathematics Department, Texas A & M University, College Station, TX 77843-3368

African Fractals: Modern Computing and Indigenous Design. By Ron Eglash. Rutgers University Press, 1999, ix + 258 pp., $25.

Reviewed by **James V. Rauff**

Ethnomathematics has enjoyed some fame and much notoriety in mathematical circles. Enthusiasts praise the recognition of mathematical ideas in non-western cultures, while critics bemoan the perceived lack of rigor and the tendency toward postmodernist excesses. Both camps are sure to cite Ron Eglash's *African Fractals* as evidence for their arguments.

Eglash has assembled a wide variety of ethnographic data, political-cultural commentary, and mathematics into an arresting tale of discovery and extrapolation. This book presents a roller-coaster ride through a broad range of academic disciplines, mostly not mathematical. To give the reader a feel for the experience of reading *African Fractals*, I will follow Eglash's organization.

The book begins with a short discussion of the nature of fractal geometry, the "five essential components" being recursion, scaling, self-similarity, infinity, and fractal dimension. Eglash includes the standard illustrative examples common to books dealing with fractals and chaotic systems: the Cantor middle-third set and the Koch curve. He uses the five essential components as metrics to help determine which of the wide range of African designs and knowledge systems are indeed instances of fractal geometry. Although the introductory chapter is not particularly compelling in itself, it nicely sets up the next chapter, the most appealing part of *African Fractals*.

Aerial photography holds a magical power for many people, myself included, and Eglash wields his overhead views of African architecture like a wizard. Pairing the photographs with computer-generated drawings, Eglash argues persuasively that there is indeed fractal geometry at work in the design of African settlements. Seeing the pictures, no one can doubt that the palace of the chief in Logone-Birni and the Ba-ila settlement are recursively generated and possess self-similarity. *African Fractals* is worth the $25 price of admission if all the reader does is gaze at these photographs and the accompanying computer-generated fractal patterns.

The architectural evidence for fractal geometry in some African cultures is convincing, but *African Fractals* goes on to suggest a deeper and more pervasive fractal character to African culture. First, however, Eglash must deal with two issues.

Is fractal geometry the result of some universal human characteristic, or does it arise only in certain cultures? In Chapter Three, Eglash points out that we do not see fractals in the settlement architecture of either Native America or Europe. This chapter is the weakest in the book, for it distracts the reader from the main flow of the discussion of African culture. Moreover, I lost some faith in

the narrative because of two glaring errors in this chapter, particularly troublesome after the outstanding images in the previous chapter. First, Eglash attributes the construction of Teotihuacan in Mexico to the Maya (p. 42). While there is considerable evidence of Teotihuacano influence and trade with the Maya world, particularly at Copan, it is not generally accepted that Teotihuacan was a Maya city; for more information, visit the Teotihuacan website `http://archaeology.la.asu.edu/teo/` maintained by Arizona State University. The second major factual error is Eglash's attribution of the mathematical analysis of Warlpiri sand drawings to Marcia Ascher in her book *Ethnomathematics* [**1**]. While Ascher does discuss Warlpiri kinship systems, she has nothing to say about the "algorithmic properties" of their sand drawings. Apparently, Eglash has confused the Warlpiri, inhabitants of continental Australia, with the Malekula, whose sand drawings Ascher does analyze. The only publications about the mathematical properties of Warlpiri sand drawings that I am aware of are the two notes [**5**] and [**7**].

The second issue that Eglash must deal with before proceeding is a central one for ethnomathematics: what constitutes mathematics in culture, and what does not? Eglash does a very nice job of delineating and classifying the possibilities, and he is careful to apply these categories to his subsequent analysis of fractal designs in African culture. Fractal designs of a group of people might be understood in different ways, ranging from unintentional by-products of some other activity to intentional designs whose character may be either implicitly or explicitly realized. This continuum is central to a raging debate in ethnomathematics. Does the existence of mathematical structures in a culture in and of itself count as mathematics, or does the mathematical creation need to be intentional? Eglash positions himself firmly in the camp that requires intentionality.

Specific examples of African designs that exhibit one or more of the "five essential components of fractal geometry" constitute the second part of *African Fractals*. Eglash also places the examples on the continuum of intentionality. The following list gives an idea of the breadth of Eglash's work.

Geometric algorithms. These include Mangbetu iterative squares sculptures, and Chokwe sand drawings that are intentionally designed as Eulerian paths.

Scaling geometry. Examples include windscreen designs showing power-law scaling, kente cloth stretching, logarithmic scaling in Ghanaian designs, and adaptive scaling in hairstyles. The discussion of windscreen designs powerfully affirms the sophistication of traditional engineering and architecture.

Numeric systems. Eglash shows that Bamana sand divination is similar to pseudorandom number generation using shift registers, and he finds self-organization in the board game Owari by viewing it as a one-dimensional cellular automaton.

Recursion. African culture is permeated by recursive design and recursive knowledge systems. We see many examples drawn from religion, dance, kinship, sculpture, and weaving. Iteration and self-reference abound. After reading this jam-packed chapter, no one will deny the potency of recursion in African thought.

Infinity. African knowledge systems use infinity "in the sense of a progression without limit" and represent it iconographically as a "completed whole." I was unsatisfied with this discussion of infinity, but, in fairness, I must say that to do the topic justice requires a book-length essay like Jadran Mimica's splendid treatment of Iqwaye counting [**4**].

Complexity. In a rapid dash through some basic notions of cybernetics and an idiosyncratic look at the Chomsky hierarchy of formal languages, Eglash argues for the existence of feedback loops and self-organization in African cultures arising from group intentionality.

The following statement summarizes the second part of the book: "four of the five basic concepts of fractal geometry—scaling, self-similarity, recursion and infinity—are all potent aspects of African mathematics," but "a quantitative measure of dimension is completely absent." Thus ends the more mathematical portion of *African Fractals*. One may quibble about some of the assertions (for instance, "mathematical complexity theory is based upon fractal geometry," which confuses chaotic dynamical systems with the theory of computation), about some omissions (the largest single source for the study of Tusona **[3]** is not even cited), and about flawed editing (Archimedean is misspelled on p. 76 and the shift-register example in Figure 7.2 has missing and incorrect entries), but the contention that fractals exist in African cultural artifacts and are intentionally and algorithmically produced is undeniably established.

Mathematicians are accustomed to books ending with suggestions for further research or applications, so they will be unprepared for the four chapters of humanistic "implications" that conclude *African Fractals*. Eglash wants us to consider what we should make of the presence of fractal geometry in African culture. He wants to get beyond the description of the fractal designs and knowledge systems and into the deeper cultural meanings. He eschews a particular epistemological framework and opts for a "toolbox" of approaches. One important tool is what he calls participant simulation, meaning the collaboration in mathematical analysis between ethnographer and informant.

Subsequent pages treat a variety of issues that are of some concern to ethnomathematicians, Africanists, and social theorists. In a discussion of the politics of recursion, Eglash states that "self-organization is not necessarily liberating; it can serve to support social control rather than resist it." He also suggests that European colonists may have failed to recognize African cities as such because the cities were organized by fractal rather than Cartesian principles. Arguing for a relation between recursion and sexuality in culture, Eglash invokes Ada Lovelace and Alan Turing. The bottom line of these arguments, as put forward by Paul Ernest **[2]** and much earlier by Raymond Wilder **[8]**, is that mathematical knowledge is a social construction. This is an important point for ethnomathematics. It must be the case that mathematical knowledge is dependent upon culture and that mathematical ideas may develop differently in different cultures before we may even consider that there is an ethnomathematics.

Eglash's assertions will appeal to those who are sympathetic to constructivist theories of mathematics, but will not convince readers unfamiliar with anthropological and cultural argumentation. His tendency to assert what some other scholar was thinking (pp. 193, 202, and 213) may be off-putting to those unaccustomed to this kind of analysis. I must also admit to being a little skeptical of the suggestion that Cantor got the idea of the middle-third set through his cousin Moritz from a design on an ancient Egyptian column (pp. 207–208).

The book closes with some observations and speculations on the future for African fractals. Eglash observes that fractal design has become part of modern African architecture. Of special interest to mathematics educators is the discussion of how knowledge of the prevalence of fractal geometry in African culture should influence how mathematics is taught. Perhaps the most important result coming

out of ethnomathematical research is that culturally informed mathematics materials are successful; see the short survey and bibliography in [**6**].

African Fractals is not a perfect book, but it is a book that mathematicians should take a look at. It is the first book to treat a single mathematical notion from the point of view of African culture, and it raises several important questions about exactly what should be and can be considered mathematics. It is a good place to see what ethnomathematics is all about, for it represents the better side of ethnomathematical research. Even if the anthropology and the mathematical and cultural philosophy seem unappealing, you should pick up a copy of *African Fractals* if only to wonder at the aerial photographs of those fractal African settlements.

REFERENCES

1. Marcia Ascher, *Ethnomathematics: A Multicultural View of Mathematical Ideas*, Brooks/Cole, Pacific Grove, CA, 1991.
2. Paul Ernest, *Social Constructivism as a Philosophy of Mathematics*, State University of New York Press, Albany, 1998.
3. Gerhard Kubik, Tusona–Luchazi ideographs: A graphic tradition practised by a people of West-Central Africa, *Acta Ethnologica et Linguistica, Series Africana* **18** (1987).
4. Jadran Mimica, *Intimations of Infinity: The Mythopoeia of the Iqwaye Counting System and Number*, Berg Publishers, Inc., Oxford, 1992.
5. James V. Rauff, Algebraic structures in Walbiri iconography, *Mathematical Connections* **1** (1993), no. 2, 5–11.
6. James V. Rauff, My brother does not have a pickup: Ethnomathematics and mathematics education, *Mathematics and Computer Education* **30** (1996), no. 1, 42–50.
7. James V. Rauff, Sand songs: the formal languages of Walpiri iconography, *Humanistic Mathematics Network Journal* **15** (1997) 17–27.
8. Raymond Wilder, *Mathematics as a Cultural System*, Pergamon Press, Oxford, 1981.

Millikin University, Decatur, IL 62522
jrauff@mail.millikin.edu

TELEGRAPHIC REVIEWS

Edited by **Arnold Ostebee**
with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges

Telegraphic Reviews are designed to alert readers in a timely manner to new books appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T : Textbook	*P : Professional Reading*	*1–4: Semester*
C : Computer Software	*L : Undergraduate Library*	*** : Special Emphasis*
S : Supplementary Reading	*13: Grade Level*	*?? : Questionable*

Readers are advised that price information is subject to change. Selected books receive a second, more extensive review in the *Monthly*.

Books submitted for review should be sent to *Book Reviews Editor, American Mathematical Monthly, St. Olaf College, 1520 St. Olaf Avenue, Northfield, MN 55057-1098.*

General, L. *You Can't Get Lost On a Golf Course.* Francis Scheid. Jones River Pr (Plymouth, MA 02360), 1998, x + 158 pp, $16.95 (P). [ISBN 1-888964-03-0] Fictional dialogs between 'student' and 'professor' on golf with mathematical digressions. Golf course topology is mostly graph theoretic. SN

General, P. *Spectral Problems in Geometry and Arithmetic.* Ed: Thomas Branson. Contemp. Math., V., 237. AMS, 1999, xi + 174 pp, $37 (P). [ISBN 0-8218-0940-7] Proceedings of a 1997 NSF-CBMS conference at the University of Iowa.

General, P. *Proceedings of the St. Petersburg Mathematical Society, Volume V.* Ed: N.N. Uraltseva. AMS Transl., Ser. 2, V. 193. AMS, 1999, vii + 271 pp, $99. [ISBN 0-8218-1390-0]

Education, P. *Mathematics in the Middle.* Ed: Larry Leutzinger. NCTM, 1998, vi + 218 pp, $26 (P). [ISBN 0-87353-460-3] 28 papers outline issues, present *Standards*-based curricula, and describe projects and activities for middle-grades classrooms.

Logic, P. *Truth, Proof and Infinity: A Theory of Constructions and Constructive Reasoning.* Peter Fletcher. Synthese Library, V. 276. Kluwer Academic, 1998, ix + 469 pp, $157. [ISBN 0-7923-5262-9] Develops precise definitions of construction and proof in constructive mathematics, and describes algorithmic underpinnings of intuitionist logic. For logicians, philosophers of mathematics, theoretical computer scientists. LB

Logic, P. *Advances in Contemporary Logic and Computer Science.* Eds: Walter A. Carnielli, Itala M.L. D'Ottaviano. Contemp. Math., V. 235. AMS, 1999, xi + 326 pp, $60 (P). [ISBN 0-8218-1364-1] Proceedings of the Eleventh Brazilian Conference on Mathematical Logic held in 1996 in Salvador, Bahia, Brazil.

Combinatorics, P, L. *Graph Classes: A Survey.* Andreas Brandstädt, Van Bang Le, Jeremy P. Spinrad. Mono. on Disc. Math. & Applic. SIAM, 1999, xi + 304 pp, $68 (P). [ISBN 0-89871-432-X] Survey of over 200 graph classes with emphasis on algorithmic point of view; results given without proof. A distillation of over a thousand cited references. SN

Number Theory, P. *Topics in Number Theory: In Honor of B. Gordon and S. Chowla.* Eds: Scott D. Ahlgren, George E. Andrews, Ken Ono. Math. & Its Applic., V. 467. Kluwer Academic, 1999, viii + 258 pp, $120. [ISBN 0-7923-5583-0] Papers from a 1997 conference at the Pennsylvania State University.

Number Theory, P. *Automorphic Forms, Automorphic Representations, and Arithmetic.* Eds: Robert S. Doran, Ze-Li Dou, George T. Gilbert. Proc. of Symposia in Pure Math., V. 66. AMS, 1999, $121 set, [ISBN 0-8218-0659-9] set. *Part 1*, xi + 278 pp; *Part 2*, xi + 330 pp. Proceedings of the NSF-CBMS Regional Conference on Euler Products and Eisenstein Series held at Texas Christian University in 1996.

Linear Algebra, T(16–17), P. *Orthogonal Sets and Polar Methods in Linear Algebra: Applications to Matrix Calculations, Systems of Equations, Inequalities, and Linear Programming.* Enrique Castillo, *et al.* Pure & Appl. Math. Wi-

ley, 1999, xii + 422 pp, $89.95. [ISBN 0-471-32889-8] Presents methods for solving a variety of linear algebra problems based on properties of orthogonal sets and dual cones. AO

Group Theory, T(18), P. *Completely Regular Semigroups.* Mario Petrich, Norman R. Reilly. Canadian Math. Soc. Ser. of Mono. & Adv. Texts, V. 23. Wiley, 1999, x + 481 pp, $115. [ISBN 0-471-19571-5]

Calculus, S(14), C. *Multivariable Calculus and Mathematics: With Applications to Geometry and Physics.* Kevin R. Coombes, Ronald L. Lipsman, Jonathan M. Rosenberg. Springer-Verlag, 1998, xiii + 283 pp, $34.95 (P), with disk. [ISBN 0-387-98360-0] Some chapters discuss standard material from an unusual point of view; others present topics that are difficult to address without a computer. Physical applications (e.g., Kepler's laws, electromagnetism, fluid flow) and elementary ideas from differential geometry play a prominent role. Exercises exploit the graphical, symbolic, and numerical capabilities of *Mathematica.* Presumes familiarity with *Mathematica* basics. AO

Calculus, T(13: 2), C. *Calculus: The Language of Change, Second Edition.* Keith D. Stroyan. Academic Pr, 1998, xix + 555 pp, with CD-ROM. [ISBN 0-12-673030-X]; *Projects for Calculus: The Language of Change, Second Edition*, 1999, xii + 326 pp, (P). [ISBN 0-12-673031-8] *First Edition* was entitled *Calculus Using Mathematica* (TR, February 1994). Text now accommodates use of either Maple or Mathematica. *Projects* book contains 52 open-ended projects on applications in mathematics, economics, science, and engineering. CD-ROM includes Mathematica Notebooks and Maple Worksheets as well as optional material not in the printed text. AO

Calculus, T(13: 2). *Calculus: Applications and Technology For Business, Social and Life Sciences, Second Edition.* Edmond C. Tomastik. Saunders College, 1999, xxiv + 533 pp, $48 (P). [ISBN 0-03-021559-5] Based on the author's text *Applied Calculus With Technology.* This edition adds material on modeling using regression analysis, includes more referenced (i.e., "real life") examples and exercises, and places greater emphasis on the use of technology. AO

Complex Analysis, P. *Convex and Starlike Mappings in Several Complex Variables.* Sheng Gong. Math. & Its Applic., V. 435. Kluwer Academic, 1998, xiii + 199 pp, $89. [ISBN 0-7923-4964-4]

Partial Differential Equations, P. *Global Solutions of Nonlinear Schrödinger Equations.* J. Bourgain. Colloquium Pub., V. 46. AMS, 1999, viii + 182 pp, $35. [ISBN 0-8218-1919-4]

Partial Differential Equations, T(17–18). *Linear and Nonlinear Waves.* G.B. Whitham. Pure & Appl. Math. Wiley, 1999, xvi + 636 pp, $64.95 (P). [ISBN 0-471-35942-4] Paperback edition of 1974 text (TR, February 1975).

Partial Differential Equations, T(15–17). *Numerical Solution of Partial Differential Equations in Science and Engineering.* Leon Lapidus, George F. Pinder. Wiley, 1999, xiv + 677 pp, (P). [ISBN 0-471-35944-0] Paperback edition of 1982 text (TR, January 1983).

Partial Differential Equations, T(17–18). *Numerical Partial Differential Equations: Conservation Laws and Elliptic Equations.* J.W. Thomas. Texts in Appl. Math., V. 33. Springer-Verlag, 1999, xxii + 556 pp, $59.95. [ISBN 0-387-98346-5] From the Preface: "... includes stability results for difference schemes for solving initial-boundary-value problems, analytic and numerical results for both scalar and systems of conservation laws, numerical methods for elliptic equations and an introduction to methods for irregularly shaped regions and for computation problems that need grid refinements."

Operator Theory, P. *Singular Quadratic Forms in Perturbation Theory.* Volodymyr Koshmanenko. Math. & Its Applic., V. 474. Kluwer Academic, 1999, viii + 308 pp, $165. [ISBN 0-7923-5625-X]

Analysis, P. *Continued Fractions: From Analytic Number Theory to Constructive Approximation.* Eds: Bruce C. Berndt, Fritz Gesztesy. Contemp. Math., V. 236. AMS, 1999, xviii + 379 pp, $78 (P). [ISBN 0-8218-1200-9] Proceedings of a 1998 conference held at the University of Missouri–Columbia.

Topology, P. *Low Dimensional Topology.* Ed: Hanna Nencka. Contemp. Math., V. 233. AMS, 1999, xi + 249 pp, $56 (P). [ISBN 0-8218-0884-2] Proceedings of a 1998 conference held in Funchal, Madeira, Portugal.

Optimization, T(17), P. *Network Optimization: Continuous and Discrete Models.* Dimitri P. Bertsekas. Athena Scientific, 1998, xiii + 593 pp, $74. [ISBN 1-886529-02-7] Introductory treatment of linear, nonlinear, and discrete network optimization problems. Theoretical results, algorithms, and applications. AO

Optimization, T(17–18), P. *Integer and Combinatorial Optimization.* George Nemhauser, Laurence Wolsey. Ser. in Disc. Math. & Optim. Wiley, 1999, xiv + 763 pp, $74.95 (P).

[ISBN 0-471-35943-2] Paperback edition of 1988 text (TR, February 1989).

Optimization, T(17–18), P. *Iterative Methods for Optimization.* C.T. Kelley. Frontiers in Appl. Math. SIAM, 1999, xv + 180 pp, $37 (P). [ISBN 0-89871-433-8] Introduction to numerical methods for unconstrained and bound-constrained optimization problems. Focuses on a few methods that convey essential ideas. Discusses deterministic sampling algorithms (e.g., Hooke-Jeeves, implicit filtering, MDS) as well as gradient-based techniques. AO

Game Theory, P. *Avances en Teoríde Juegos con Aplicaciones Económicas Y Sociales.* Eds: J.M. Bilbao, F.R. Fernández. Universidad de Sevilla (Porvenir, 27–41013 Sevilla, Spain), 1999, ix + 254 pp, (P). [ISBN 84-472-0517-7]

Mathematical Modeling, T(15–17). *Mathematical Modeling, Second Edition.* Mark M. Meerschaert. Academic Pr, 1999, xvi + 351 pp. [ISBN 0-12-487652-8] New sections on discrete modeling (including integer programming) and on chaos and fractals. Three parts discuss optimization models, dynamical models, and probability models. (*First Edition*, TR, December 1993.) AO

Optimal Control, P. *Nonlinear Analysis, Differential Equations and Control.* Eds: F.H. Clarke, R.J. Stern. NATO Sci. Ser. C, V. 528. Kluwer Academic, 1999, xix + 602 pp, $297. [ISBN 0-7923-5665-9] Proceedings of a 1998 NATO Advanced Study Institute held in Montréal, Canada.

Optimal Control, T(16–17: 1, 2), S, P. *Fuzzy Control of Industrial Systems: Theory and Applications.* Ian S. Shaw. Kluwer Academic, 1998, xxiii + 192 pp, $115. [ISBN 0-7923-8249-8] Includes background on fuzzy sets and fuzzy logic, but assumes familiarity with linear continuous and discrete systems, feedback, basic control theory. Covers fuzzy controllers, rule-based systems, stability analysis, neurofuzzy controllers, practical development issues, collaborative control systems (fuzzy controller plus human operator). No exercises in text, but questions over each chapter are available on request from the author. LB

Stochastic Processes, P. *Advances in Stochastic Inequalities.* Eds: Theodore P. Hill, Christian Houdré. Contemp. Math., V. 234. AMS, 1999, ix + 212 pp, $45 (P). [ISBN 0-8218-1086-3] Papers from a Special Session at the AMS meeting held at the Georgia Institute of Technology in 1997.

Stochastic Processes, T(16–17), P. *Stochastic Processes.* Emanuel Parzen. Classics in Appl. Math., V. 24. SIAM, 1999, xix + 324 pp, $37 (P). [ISBN 0-89871-441-9] Unabridged, corrected republication. Originally published in 1962 by Holden-Day.

Stochastic Processes, P. *Limit Theorems for Random Fields with Singular Spectrum.* Nikolai Leonenko. Math. & Its Applic., V. 465. Kluwer Academic, 1999, viii + 401 pp, $192. [ISBN 0-7923-5635-7]

Elementary Statistics, T(13: 1), C. *Statistics: An Introduction, Fifth Edition.* Robert D. Mason, Douglas A. Lind, William G. Marchal. Duxbury Pr, 1998, xxii + 689 pp, $79.95, with disk. [ISBN 0-534-35379-7] Added in this edition: Excel spreadsheet examples, a chapter on quality control, and a section on box plots. (*Fourth Edition*, TR, November 1994.) KB

Elementary Statistics, T(14: 1). *Applied Statistics and Probability for Engineers, Second Edition.* Douglas C. Montgomery, George C. Runger. Wiley, 1999, xiv + 917 pp, $103.95. [ISBN 0-471-17027-5] Stresses methodology and applications rather than mathematical theory. Uses output from modern statistical software in examples. All text data are available in electronic form. HS

Statistical Methods, T(17–18: 1), P, L. *Model Selection and Inference: A Practical Information-Theoretic Approach.* Kenneth P. Burnham, David R. Anderson. Springer-Verlag, 1998, xx + 353 pp, $69.95. [ISBN 0-387-98504-2] Introduction to analysis of empirical data using information-theoretic approaches. Presents a consistent methodology that treats model formulation, model selection, estimation of model parameters, and their uncertainty in a unified manner. Reviews other information criteria (e.g., AIC_c, QAIC, TIC) and presents several new approaches to estimating model selection uncertainty and incorporating this uncertainty into estimates of precision. Examples (primarily ecological) illustrate technical issues. KB

Statistical Methods, S(15–17), P. *Selecting and Ordering Populations: A New Statistical Methodology.* Jean Dickinson Gibbons, Ingram Olkin, Milton Sobel. Classics in Appl. Math., V. 26. SIAM, 1999, xxv + 569 pp, $59.50 (P). [ISBN 0-89871-439-7] Unabridged, corrected republication. (1977 Wiley edition, TR, April 1978.)

Statistical Methods, T(18), P. *Fractional Factorial Plans.* Aloke Dey, Rahul Mukerjee. Ser. in Prob. & Stat. Wiley, 1999, xii + 211 pp, $84.95. [ISBN 0-471-29414-4]

Statistics, S(18), P. *Geostatistics: Modeling*

Spatial Uncertainty. Jean-Paul Chilès, Pierre Delfiner. Ser. in Prob. & Stat. Wiley, 1999, xi + 695 pp, $125. [ISBN 0-471-08315-1]

Theory of Computation, T(15–17: 1), S, P. *Term Rewriting and All That.* Franz Baader, Tobias Nipkow. Cambridge Univ Pr, 1998, xii + 301 pp, $49.95. [ISBN 0-521-45520-0] Discusses a branch of theoretical computer science based on equational logic which uses equations as directed replacement rules. Requires only basic discrete math background. Applications to algebra (word problem), software engineering, recursion theory, functional and logic programming. LB

Applications (Economics), P. *Financial Markets: Stochastic Analysis and the Pricing of Derivative Securities.* A.V. Mel'nikov. Transl. of Math. Mono., V. 184. AMS, 1999, xiv + 133 pp, $49. [ISBN 0-8218-1082-0]

Applications (Social Science), S(16–17), P, L. *Evaluation and Optimization of Electoral Systems.* Pietro Grilli di Cortona, *et al.* Mono. on Disc. Math. & Applic. SIAM, 1999, xvi + 230 pp, $53 (P). [ISBN 0-89871-422-2] Presents a quantitative approach to the study of electoral systems based on single- and multiple-criteria optimization methods. Also discusses automatic procedures for designing electoral districts. Includes reflections on the political aspects of election reform that provide perspective on the benefits and the limits of using mathematical models to design electoral systems. AO

Applications (Systems Theory), T(16–18), P. *Classical Control Using H^∞ Methods: Theory, Optimization, and Design.* J. William Helton, Orlando Merino. SIAM, 1998, xvi + 292 pp, $45 (P). [ISBN 0-89871-419-2] Parts I–II provide an introduction to control systems design; Parts III–V develop a theory for frequency domain CAD. Assumes reader has some background in control theory (e.g., from a one-semester introductory course).

Applications (Systems Theory), S(16–18). *Classical Control Using H^∞ Methods: An Introduction to Design.* J. William Helton, Orlando Merino. SIAM, 1998, xii + 171 pp, $18.50 (P). [ISBN 0-89871-424-9] Contains the design portions of the author's text *Classical Control Using H^∞ Methods: Theory, Optimization, and Design.* (See preceding review.)

Applications (Systems Theory), T(18), P. *Indefinite-Quadratic Estimation and Control: A Unified Approach to H^2 and H^∞ Theories.* Babak Hassibi, Ali H. Sayed, Thomas Kailath. Stud. in Appl. & Num. Math., V. 16. SIAM, 1999, xvii + 555 pp, $80. [ISBN 0-89871-411-7]

Applications (Systems Theory), P. *Lecture Notes in Control and Information Sciences–245: Robustness in Identification and Control.* Eds: Andrea Garulli, Alberto Tesi, Antonio Vicino. Springer-Verlag, 1999, xiv + 467 pp, $105 (P). [ISBN 1-85233-179-8] Papers, many in a tutorial style, from a 1998 workshop held in Siena, Italy.

Applications (Systems Theory), P. *Lecture Notes in Control and Information Sciences–243: Progress in System and Robot Analysis and Control Design.* Eds: S.G. Tzafestas, G. Schmidt. Springer-Verlag, 1999, xxvii + 588 pp, $125 (P). [ISBN 1-85233-123-2] Papers from the Third European Robotics, Intelligent Systems, and Control Conference (EURISCON '98) held in Athens, Greece.

Applications (Systems Theory), P. *Lecture Notes in Control and Information Sciences–244: New Directions in Nonlinear Observer Design.* Eds: H. Nijmeijer, T.I. Fossen. Springer-Verlag, 1999, xix + 525 pp, $119 (P). [ISBN 1-85233-134-8] Papers from a 1999 workshop held in Geiranger Fjord, Norway.

Applications, T(16–17: 1), L. *NURBS: From Projective Geometry to Practical Use, Second Edition.* Gerald E. Farin. AK Peters, 1999, xv + 267 pp, $44. [ISBN 1-56881-084-9] NURBS (Non-Uniform Rational B-Splines) are widely used for describing objects in CAD/CAM and computer graphics applications. Text develops concepts and algorithms for describing curves and surfaces using NURBS. *Second Edition* incorporates recent research results and a new chapter on Pythagorean curves (*First Edition*, TR, November 1995). AO

Applications, P, L. *Computation, Causation, and Discovery.* Eds: Clark Glymour, Gregory F. Cooper. MIT Pr, 1999, xvi + 552 pp, $45 (P). [ISBN 0-262-57124-2] A collection of papers on data mining for causal relationships. (For example, to answer the question "Does an intervention that directly affects one part of a system indirectly alter another part?") In five parts: Causation, Representation and Prediction; Search; Controversy Over Search; Estimating Causal Effects; Scientific Applications.

Applications, P. *Modern Mathematical, Management, & Statistical Sciences, I.* Eds: Edward J. Dudewicz, Bruce L. Golden, Z. Govindarajulu. Amer. J. of Math. & Manag. Sci., V. 18, Nos. 3 & 4. American Sciences Pr, 1998, 189 pp, $195 (P). [ISBN 0-935950-45-1]

Reviewers

KB: Karla Ballman, Macalester; LB: Lynne Baur, Carleton; SN: Sam Northshield, Carleton; AO: Arnold Ostebee, St. Olaf; HS: Heidi Shierholz, St. Olaf.

UNITED STATES POSTAL SERVICE™

Statement of Ownership, Management, and Circulation

(Required by 39 USC 3685)

1. Publication Title: The American Mathematical Monthly
2. Publication Number: 0 0 0 2 - 9 8 9 0
3. Filing Date: 9/14/99
4. Issue Frequency: Monthly, except bimonthly June/July, Aug./Sept
5. Number of Issues Published Annually: 10
6. Annual Subscription Price: $20
7. Complete Mailing Address of Known Office of Publication *(Not printer) (Street, city, county, state, and ZIP+4)*: MAA, 1529 18th St. Washington, D.C. 20036
   Contact Person: H. Waldman
   Telephone: 202-387-5200
8. Complete Mailing Address of Headquarters or General Business Office of Publisher *(Not printer)*: same
9. Full Names and Complete Mailing Addresses of Publisher, Editor, and Managing Editor *(Do not leave blank)*
   Publisher *(Name and complete mailing address)*: MAA, 1529 18th St. Washington, D. C. 20036-1385
   Editor *(Name and complete mailing address)*: Roger Horn, University of Utah, Salt Lake City, UT 84112
   Managing Editor *(Name and complete mailing address)*: Harry Waldman
10. Owner *(Do not leave blank. If the publication is owned by a corporation, give the name and address of the corporation immediately followed by the names and addresses of all stockholders owning or holding 1 percent or more of the total amount of stock. If not owned by a corporation, give the names and addresses of the individual owners. If owned by a partnership or other unincorporated firm, give its name and address as well as those of each individual owner. If the publication is published by a nonprofit organization, give its name and address.)*

Full Name	Complete Mailing Address
The Mathematical Association of America	1529 18th St. Washington, D.C. 20036

11. Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages, or Other Securities. If none, check box → ☒ None

Full Name	Complete Mailing Address

12. Tax Status *(For completion by nonprofit organizations authorized to mail at special rates) (Check one)*
    The purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes:
    ☒ Has Not Changed During Preceding 12 Months
    ☐ Has Changed During Preceding 12 Months *(Publisher must submit explanation of change with this statement)*

PS Form 3526, September 1995 *(See Instructions on Reverse)*

13. Publication Title: The American Mathematical Monthly
14. Issue Date for Circulation Data Below: Aug/Sept 1999

15. Extent and Nature of Circulation		Average No. Copies Each Issue During Preceding 12 Months	Actual No. Copies of Single Issue Published Nearest to Filing Date
a. Total Number of Copies *(Net press run)*		18,000	18,000
b. Paid and/or Requested Circulation	(1) Sales Through Dealers and Carriers, Street Vendors, and Counter Sales *(Not mailed)*	0	0
	(2) Paid or Requested Mail Subscriptions *(Include advertiser's proof copies and exchange copies)*	15,500	15,500
c. Total Paid and/or Requested Circulation *(Sum of 15b(1) and 15b(2))*		15,500	15,500
d. Free Distribution by Mail *(Samples, complimentary, and other free)*		2,50 0	2,500
e. Free Distribution Outside the Mail *(Carriers or other means)*		0	0
f. Total Free Distribution *(Sum of 15d and 15e)*		2,500	2,500
g. Total Distribution *(Sum of 15c and 15f)*		18,000	18,000
h. Copies not Distributed	(1) Office Use, Leftovers, Spoiled	0	0
	(2) Returns from News Agents	0	0
i. Total *(Sum of 15g, 15h(1), and 15h(2))*		18,000	18,000
Percent Paid and/or Requested Circulation *(15c / 15g x 100)*		86%	86%

16. Publication of Statement of Ownership
    ☒ Publication required. Will be printed in the November issue of this publication.
    ☐ Publication not required.
17. Signature and Title of Editor, Publisher, Business Manager, or Owner: Donald J. Albers, Director of Publications — Date: 9/14/99

I certify that all information furnished on this form is true and complete. I understand that anyone who furnishes false or misleading information on this form or who omits material or information requested on the form may be subject to criminal sanctions (including fines and imprisonment) and/or civil sanctions (including multiple damages and civil penalties).

Instructions to Publishers

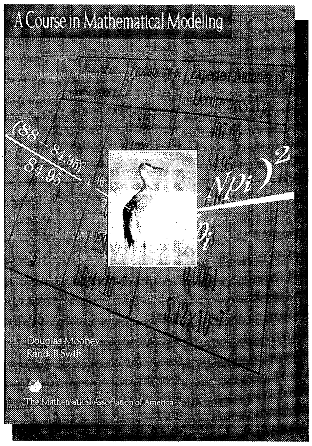
1. Complete and file one copy of this form with your postmaster annually on or before October 1. Keep a copy of the completed form for your records.
2. In cases where the stockholder or security holder is a trustee, include in items 10 and 11 the name of the person or corporation for whom the trustee is acting. Also include the names and addresses of individuals who are stockholders who own or hold 1 percent or more of the total amount of bonds, mortgages, or other securities of the publishing corporation. In item 11, if none, check the box. Use blank sheets if more space is required.
3. Be sure to furnish all circulation information called for in item 15. Free circulation must be shown in items 15d, e, and f.
4. If the publication had second-class authorization as a general or requester publication, this Statement of Ownership, Management, and Circulation must be published; it must be printed in any issue in October or, if the publication is not published during October, the first issue printed after October.
5. In item 16, indicate the date of the issue in which this Statement of Ownership will be published.
6. Item 17 must be signed.

Failure to file or publish a statement of ownership may lead to suspension of second-class authorization.

The Mathematical Association of America

A Course in Mathematical Modeling

Douglas Mooney and Randall Swift



Series: Classroom Resource Materials

A Course in Mathematical Modeling is intended as a text for a modeling course accessible to students who have mastered a one year course in calculus. Mooney and Swift's approach to modeling is presented balancing theoretical versus empirical models, analytic models versus simulation. deterministic versus stochastic models, and discrete versus continuous models. Most examples are drawn from real world data or from models that have been used in various applied fields. The use of computers in both simulation and analysis is an integral part of the presentation.

The authors emphasize teaching modeling as opposed to presenting models. beginning their book with the simple discrete exponential growth model as a building block, and successively refining it. This refinement includes adding variable growth rates and multiple variables, fitting growth rates to data, including random elements, testing goodness of fit, using computer simulations, and moving to a continuous setting.

Students taking a course based on this book should have some mathematical maturity, but will need little advanced knowledge. The book presents more advanced topics on an as needed basis and serves to show how the different topics of undergraduate mathematics can be used together to solve problems. This perspective is valuable as either a road map for the beginning student or as a capstone for the more advanced students. The course presents elements of discrete dynamical systems, basic probability theory, differential equations, matrix algebra, stochastic processes, curve fitting, statistical testing, and regression analysis. Computer analysis is extensively used in conjunction with these topics.

You can also use this book if you are seeking applications to supplement a course in linear algebra, differential equations, difference equations, probability theory or statistics

Catalog Code: MML/JR 500 pp., Paperbound, 1999 ISBN 0-88385-712-x List: $41.95 MAA Member: $32.95

Name__________ Credit Card No.__________

Address__________ Signature__________ Exp. Date____/____

City__________ Qty_____ Price $_____ Amount $_____

State__________ Zip __________ Shipping and Handling $_____

Phone __________ **Catalog Code: MML/JR** Total $_____

Shipping and Handling: Postage and handling are charged as follows: USA orders (shipped via UPS): $2.95 for the first book, and $1.00 for each additional book. **Canadian orders:** $4.50 for the first book and $1.50 for each additional book. Canadian orders will be shipped within 10 days of receipt of order via the fastest available route. We do not ship via UPS into Canada unless the customer specially requests this service. Canadian customers who request UPS shipment will be billed an additional 7% of their total order. **Overseas Orders:** $3.50 per item ordered for books sent surface mail. Airmail service is available at a rate of $7.00 per book. Foreign orders must be paid in US dollars through a US bank or through a New York clearinghouse. **Credit card orders are accepted for all customers. All orders must be prepaid with the exception of books purchased for resale by bookstores and wholesalers.**

Order Via:

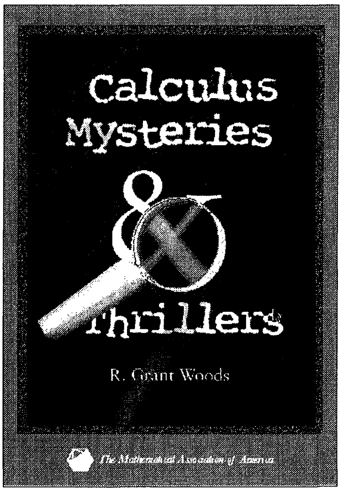
Phone: **1 (800) 331.1622**

Fax: **(301) 206.9789**

Mail: **Mathematical Association of America**
PO Box 91112
Washington, DC 20090-1112

Web: **www.maa.org**

THE MATHEMATICAL ASSOCIATION OF AMERICA



Calculus Mysteries and Thrillers

R. Grant Woods

Series: Classroom Resource Materials

Text Supplement

Calculus projects you can give to your students . . .

This book consists of eleven mathematics projects based on introductory single-variable calculus, together with some guidance on how to make use of them. Each project is presented as an amusing short story. In many of them a group of undergraduate mathematics students, formed into a consulting company called *Math Iz Us*, is hired to solve mathematical problems brought to them by clients. The problems solved include: helping to prosecute an accused pool shark, defending a driver accused of speeding, assisting a hockey coach in making his star forward a more effective goal scorer, and advising a pirate captain on how to divide a gold-plated goose-egg fairly among his crew.

In each problem, the problem solvers are required to present to their client a detailed written report of their findings. Thus, students must produce and analyze accurate mathematical models of complex, verbally presented "real life" situations, and write a clear technical account of their solution.

Instructors who are looking for problems that are novel, interesting, and several levels more complex than the typical text book "word problem" will find them in this book. It will be of particular value to instructors who wish to combine training in applications of calculus with training in technical writing. The complexity of the problems makes them suitable for use as group projects.

The calculus concepts on which the problems are based include: tangent and normal lines, optimization by use of critical points, inverse trig functions, volumes of solids, surface area integrals, and modeling economic concepts using definite integrals. Although a few ideas from physics and economics are used in the problems, no prior knowledge of these fields is required.

Catalog Code: CTM/JR

144 pp., Paperbound, 1998, ISBN 0-88385-711-1

List: $24.95 MAA Member: $19.95

Phone in Your Order Now! ☎ *1-800-331-1622*

Monday – Friday 8:30 am – 5:00 pm FAX (301) 206-9789

or mail to: The Mathematical Association of America, PO Box 91112, Washington, DC 20090-1112

Shipping and Handling: Postage and handling are charged as follows: **USA orders (shipped via UPS):** $2.95 for the first book, and $1.00 for each additional book. **Canadian orders:** $4.50 for the first book and $1.50 for each additional book. Canadian orders will be shipped within 10 days of receipt of order via the fastest available route. We do not ship via UPS into Canada unless the customer specially requests this service. Canadian customers who request UPS shipment will be billed an additional 7% of their total order. **Overseas orders:** $3.50 per item ordered for books sent surface mail. Airmail service is available at a rate of $7.00 per book. Foreign orders must be paid in US dollars through a US bank or through a New York clearinghouse. Credit Card orders are accepted for all customers.

Name ______________________

Address ______________________

City ______________ State _____ Zip ________

Phone ______________________

QTY.	CATALOG CODE	PRICE	AMOUNT
_____	CTM/JR	_____	_____

All orders must be prepaid with the exception of books purchased for resale by bookstores and wholesalers.

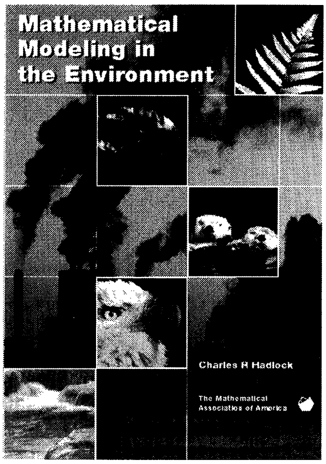
Shipping & handling ________

TOTAL ________

Payment ☐ Check ☐ VISA ☐ MasterCard

Credit Card No. ______________ Expires ___/___

Signature ______________________



THE MATHEMATICAL ASSOCIATION OF AMERICA

Mathematical Modeling in the Environment

Charles Hadlock

Series: Classroom Resource Materials

Packaged with a PC compatible disk that enhances the material in the text.

Suitable for classroom adoption in an innovative course for

- **a general education mathematics elective**
- **a mathematics or science major advanced elective**
- **an interdisciplinary course, even at a relatively elementary level**
- **a mathematical modeling course in a civil/environmental engineering program**

This book has a dual objective: first, to introduce the reader to some of the most important and widespread environmental issues of the day; and second, to illustrate the vital role played by mathematical models in investigating these issues. The environmental issues addressed include: ground-water contamination, air pollution, and hazardous material emergencies. These issues are presented in their full real-world context, not as scientific or mathematical abstractions; and for background, readers are invited to investigate their status in their own communities.

The first part of the book leads the reader through relatively elementary modeling of these phenomena, including simple algebraic equations for ground water, slightly more complex algebraic equations (preferably implemented on a spreadsheet or other computerized framework) for air pollution, and a fully computerized modeling package for hazardous materials incident analysis. The interplay between physical intuition and mathematical analysis is emphasized.

For more advanced readers, the second part of the book returns to the same three subjects but with a higher level of mathematical sophistication (adjustable to the preparation of the reader by selection of subsections.) Many important classical mathematical themes are developed through this context, examples coming from single and multivariable calculus, differential equations, numerical analysis, linear algebra and probability. The material is presented in such a way as to minimize the required background and to encourage the subsequent study of some of these fields.

An elementary course for a general audience could be based entirely on Part I, and a higher level mathematics, science, or engineering course could move quickly to Part 2.

A PC compatible diskette packaged with the text contains a spreadsheet program that facilitates the numerical experimentation with the Gaussian plume equation introduced in Chapter 3, as well as public domain DOS program (ARCHIE) for evaluating the consequences from various hazardous materials scenarios (e.g., the physical extent of flammable and toxic vapor clouds). Text is not tied to the use of this software, but it is included as an aid to meet the pedagogical objectives of the text.

Catalog Code: ENV/SA
312 pp., Paperbound, 1998, ISBN 0-88385-709-X
List: $55.00 MAA Member: $43.95

Instructor's and Solutions Manual for *Mathematical Modeling in the Environment*

Charles Hadlock

Contains the complete solutions and further discussion of nearly every exercise presented in the textbook. This includes both the mathematical/computational exercises as well as the research questions and investigations. Readers will benefit greatly from perusing solutions to the problems whether they have worked them out themselves or not. Students using this volume will still need to work out solutions of research questions using their own sources and adapting them to their own geographic locations, or using their own computational schemes, so this volume could well be useful for students in many course contexts. Enrichment material is included on the topics of some of the exercises. Advice for teachers who lack previous environmental experience, but who want to teach this material is also provided and makes it practical for such persons to offer a course based on these volumes.

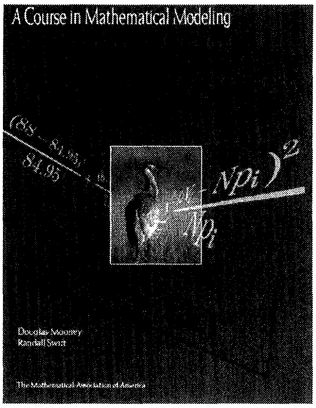
Catalog Code: EVS/SA
150 pp., Paperbound, 1998, ISBN 0-88385-713-8
List: $18.95 MAA Member: $14.95

Phone in Your Order Now! ☎ *1-800-331-1622*

Monday – Friday 8:30 am – 5:00 pm FAX (301) 206-9789

or mail to: The Mathematical Association of America, PO Box 91112, Washington, DC 20090-1112

THE MATHEMATICAL ASSOCIATION OF AMERICA



A Course in Mathematical Modeling

Series: Classroom Resource Materials

Douglas Mooney and Randall Swift

Textbook

This book is intended as a text for a modeling course accessible to students who have mastered a one year course in calculus. It balances a variety of opposing modeling methodologies including theoretical models versus empirical models, analytical models versus simulation, deterministic models versus stochastic models, and discrete models versus continuous models. Most of the examples are drawn from real-world data or from models that have been used in various applied fields. The use of computers in both simulation and in mathematical analysis is an integral part of the presentation.

The authors emphasize the teaching of the modeling process as opposed to merely presenting models. They begin their book with the simple discrete exponential growth model, and successively refine it to include variable growth rates, multiple variables, growth rates fitted to data, and the effects of random factors. The last part of the book moves into continuous-time models. Issues of model validity and purpose are emphasized throughout.

Students taking a course based on this book should have some mathematical maturity, but will need little advanced knowledge. The book presents more advanced topics on an as-needed basis and serves to show how the different topics of undergraduate mathematics can be used together to solve problems. This perspective is valuable as either a road map for beginning students or as a capstone for more advanced students. The course presents elements of discrete dynamical systems, basic probability theory, differential equations, matrix algebra, stochastic processes, curve fitting, statistical testing, and regression analysis. Computer analysis is extensively used in conjunction with these topics.

You can also use this book if you are seeking applications to supplement a course in linear algebra, differential equations, difference equations, probability theory, or statistics.

Catalog Code: MML/JR
400 pp., Paperbound, 1999
ISBN 0-88385-712-X
List: \$41.95 MAA Member: \$32.95

Visit <www.wku.edu/~swiftrj/Modeling/ modeling.html> where you can visit the authors' website and download data sets, *Mathematica* files, and other modeling resources that execute the models described in the text.

Phone in Your Order Now! ☎ *1-800-331-1622*

Monday – Friday 8:30 am – 5:00 pm FAX (301) 206-9789

or mail to: The Mathematical Association of America, PO Box 91112, Washington, DC 20090-1112

Shipping and Handling: Postage and handling are charged as follows: **USA orders (shipped via UPS):** \$2.95 for the first book, and \$1.00 for each additional book. **Canadian orders:** \$4.50 for the first book and \$1.50 for each additional book. Canadian orders will be shipped within 10 days of receipt of order via the fastest available route. We do not ship via UPS into Canada unless the customer specially requests this service. Canadian customers who request UPS shipment will be billed an additional 7% of their total order. **Overseas orders:** \$3.50 per item ordered for books sent surface mail. Airmail service is available at a rate of \$7.00 per book. Foreign orders must be paid in US dollars through a US bank or through a New York clearinghouse. Credit Card orders are accepted for all customers.

Name ______________________

Address ______________________

City ____________ State _____ Zip _______

Phone ______________________

Qty.	Catalog Code	Price	Amount
_______	MML/JR	_______	_______

All orders must be prepaid with the exception of books purchased for resale by bookstores and wholesalers.

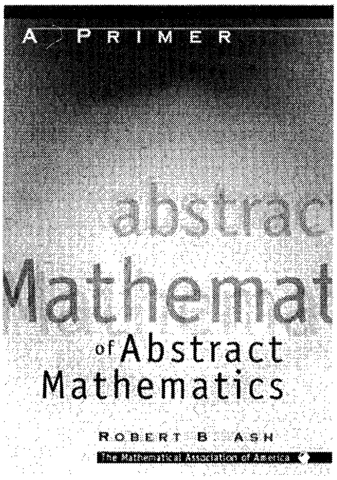
Shipping & handling _______

TOTAL _______

Payment ☐ Check ☐ VISA ☐ MasterCard

Credit Card No. ______________ Expires ___/___

Signature ______________________



THE MATHEMATICAL ASSOCIATION OF AMERICA

A Primer of Abstract Mathematics

Robert Ash

Series: Classroom Resource Materials

A Primer of Abstract Mathematics prepares the reader to cope with abstract mathematics, specifically abstract algebra. It can serve as a text for prospective mathematics majors, as well as for those students taking or preparing to take a first course in abstract algebra, or those in applied fields who need experience in dealing with abstract mathematical ideas.

Learning any area of abstract mathematics involves writing formal proofs, but it is equally important to think intuitively about the subject and to express ideas clearly and cogently. The author aids intuition by keeping proofs short and as informal as possible, using concrete examples which illustrate all the features of the general case, and by giving heuristic arguments when a formal development would take too long. The text can serve as a model on how to write mathematics for an audience with limited experience in formalism and abstraction.

Ash introduces several expository innovations in *A Primer of Abstract Mathematics*. He presents an entirely informal development of set theory that gives students the basic results that they will need in algebra. The chapter which presents the theory of linear operators introduces the Jordan Canonical Form right at the beginning, with a proof of existence at the end of the chapter.

Contents:

Logic and Foundations: Truth Tables, Quantifiers, Proofs, Sets, Functions, Relations.

Counting: Fundamentals, The Binomial and Multinomial Theorems, The Principle of Inclusion and Exclusion, Counting Infinite Sets.

Elementary Number Theory: The Euclidean Algorithm, Unique Factorization, Algebraic Structures, Further Properties of Congruence Modulo m, Linear Diophantine Equations and Simultaneous Congruences, Theorems of Euler and Fermat, The Möbius Inversion Formula.

Some Highly Informal Set Theory: Well-Orderings, Zorn's Lemma and the Axiom of Choice, Cardinal Numbers, Addition and Multiplication of Cardinals.

Linear Algebra: Matrices, Determinants and Inverses, The Vector Space F^n, Linear Independence and Bases, Subspaces, Linear Transformations, Inner Product Spaces, Eigenvalues and Eigenvectors.

Theory of Linear Operators: Jordan Canonical Form, The Minimal and Characteristic Polynomials, The Adjoint of a Linear Operator, Normal Operators, The Existence of the Jordan Canonical Form.

Appendix: An Application of Linear Algebra.

Catalog Code: BMA/JR
188 pp., Paperbound, 1998
ISBN 0-88385-709-X
List: $27.95   MAA Member: $21.95

Phone in Your Order Now! ☎ *1-800-331-1622*

Monday – Friday 8:30 am – 5:00 pm   FAX (301) 206-9789

or mail to: The Mathematical Association of America, PO Box 91112, Washington, DC 20090-1112

Shipping and Handling: Postage and handling are charged as follows: **USA orders (shipped via UPS):** $2.95 for the first book, and $1.00 for each additional book. **Canadian orders:** $4.50 for the first book and $1.50 for each additional book. Canadian orders will be shipped within 10 days of receipt of order via the fastest available route. We do not ship via UPS into Canada unless the customer specially requests this service. Canadian customers who request UPS shipment will be billed an additional 7% of their total order. **Overseas orders:** $3.50 per item ordered for books sent surface mail. Airmail service is available at a rate of $7.00 per book. Foreign orders must be paid in US dollars through a US bank or through a New York clearinghouse. Credit Card orders are accepted for all customers.

Name ______________________

Address ______________________

City ______________ State _____ Zip ________

Phone ______________________

QTY.	CATALOG CODE	PRICE	AMOUNT
______	BMA/JR	______	______

All orders must be prepaid with the exception of books purchased for resale by bookstores and wholesalers.

Shipping & handling ______

TOTAL ______

Payment ☐ Check ☐ VISA ☐ MasterCard

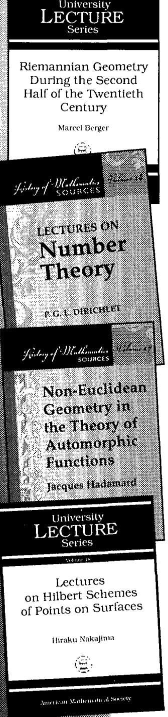
Credit Card No. ______________ Expires ___/___

Signature ______________________

American Mathematical Society

Recently Published Titles from the AMS

These selections from the History of Mathematics *Sources* series and University Lecture Series enhance the classroom experience. Consider recommending these texts to students for supplementary reading or independent study. For more AMS titles to broaden the learning experience, visit the AMS Bookstore at **www.ams.org/bookstore/**.



Supplementary Reading

Riemannian Geometry During the Second Half of the Twentieth Century

Marcel Berger, *Institut des Hautes Études Scientifiques, Bures-sur-Yvette, France*

Masterfully written and delightful to read. In addition to the numerous digressions for newly introduced concepts, the author adds to the value of the survey by providing fertile opinions, some of them his, others those of his close colleagues and of M. Gromov in particular. The wonderful effort of the author is shown partially by the long bibliography of thirty pages, with references updated right to the very end of the century. A person who wants to learn more about Riemannian geometry will certainly do him/herself a good service by reading Berger's work.

—Mathematical Reviews

Reprint arranged with the approval of the publisher B. G. Teubner, Stuttgart and Leipzig.

University Lecture Series; 1999; approximately 217 pages; Softcover; ISBN 0-8218-2052-4; List $34; All AMS members $27; Order code ULECT-BERGERMM911

A Classic

Lectures on Number Theory

P. G. L. Dirichlet with supplements by **R. Dedekind**

A new edition of Dirichlet's Lectures on Number Theory *would be big news any day, but it's particularly gratifying to see the book appear as "the first of an informal sequence" which is to include "classical mathematical works that served as cornerstones for modern mathematical thought." So all power to the American Mathematical Society and the London Mathematical Society in their joint-venture* History of Mathematics *series: May the "Sources" subseries live long and prosper. [T]his is quite accessible, and could almost be used as a textbook still today. For those who like to heed Abel's admonition to "read the masters, not their students," here's a great opportunity to learn more about Number Theory.*

—MAA Online

Co-published with the London Mathematical Society. Members of the LMS may order directly from the AMS at the AMS member price. The LMS is registered with the Charity Commissioners.

History of Mathematics, Volume 16; 1999; 275 pages; Softcover; ISBN 0-8218-2017-6; List $49; All AMS members $39; Order code HMATH/16MM911

Supplementary Reading

Non-Euclidean Geometry in the Theory of Automorphic Functions

Jacques Hadamard, and **Jeremy J. Gray**, *Open University, Milton Keynes, UK*, and **Abe Shenitzer,** *York University, Toronto, ON, Canada,* Editors

This is the English translation of a volume originally published only in Russian and now out of print. The book was written by Jacques Hadamard on the work of Poincaré.

Hadamard begins with hyperbolic geometry, which he compares with plane and spherical geometry. In Chapter 2 he presents the appropriate automorphic functions, in particular, Fuchsian functions. Chapter 4 is devoted to the outlines of the more complicated Kleinian case. Chapter 5 discusses algebraic functions and linear algebraic differential equations, and the last chapter sketches the theory of Fuchsian groups and geodesics.

This book is the second in an informal sequence of works called "History of Mathematics, Sources", to be included within the History of Mathematics series, co-published by the AMS and the London Mathematical Society.

Members of the LMS may order directly from the AMS at the AMS member price. The LMS is registered with the Charity Commissioners.

History of Mathematics, Volume 17; 1999; 95 pages; Softcover; ISBN 0-8218-2030-3; List $19; All AMS members $15; Order code HMATH/17MM911

Supplementary Reading

Lectures on Hilbert Schemes of Points on Surfaces

Hiraku Nakajima, *Kyoto University, Japan*

The Hilbert scheme $X^{[n]}$ of a surface X describes collections of n (not necessarily distinct) points on X. More precisely, it is the moduli space for 0-dimensional subschemes of X of length n. Recently it was realized that Hilbert schemes originally studied in algebraic geometry are closely related to several branches of mathematics, such as singularities, symplectic geometry, representation theory–even theoretical physics. The discussion in the book reflects this feature of Hilbert schemes.

For example, a construction of the representation of the infinite dimensional Heisenberg algebra (i.e., Fock space) is presented. This representation has been studied extensively in the literature in connection with affine Lie algebras, conformal field theory, etc. However, the construction presented in this volume is completely unique and provides the unexplored link between geometry and representation theory.

University Lecture Series, Volume 18; 1999; 132 pages; Softcover; ISBN 0-8218-1956-9; List $21; All AMS members $17; Order code ULECT/18MM911

All prices subject to change. Charges for delivery are $3.00 per order. For optional air delivery outside of the continental U. S., please include $6.50 per item. *Prepayment required.* Order from: **American Mathematical Society**, P. O. Box 5904, Boston, MA 02206-5904, USA. For credit card orders, fax 1-401-455-4046 or call toll free 1-800-321-4AMS (4267) in the U. S. and Canada, 1-401-455-4000 worldwide. Or place your order through the AMS bookstore at www.ams.org/bookstore/. Residents of Canada, please include 7% GST.

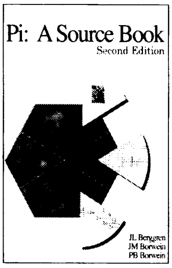
AMS
American Mathematical Society

SPRINGER FOR MATHEMATICS

LENNART BERGGREN, JONATHAN BORWEIN, and **PETER BORWEIN,** all, Simon Fraser University, Burnaby, BC, Canada

PI: A SOURCE BOOK

Second Edition



The aim of this book is to provide a complete history of π from the dawn of mathematical time to the present. The story of π reflects the most seminal, the most serious, and sometimes the silliest aspects of mathematics. One of the beauties of the literature on π is that it allows for the inclusion of very modern, yet still accessible, mathematics. Mathematicians and historians of mathematics will find this book indespensable.

1999/APP. 752 PP., 82 ILLUS./HARDCOVER/$64.95 (TENT.)
ISBN 0-387-98946-3

HENRI COHEN, Université de Bordeaux, France

ADVANCED TOPICS IN COMPUTATIONAL NUMBER THEORY

Contents: Fundamental Results and Algorithms in Dedekind domains • Basic Relative Number Field Algorithms • The Fundamental Theorems of Global Class Field Theory • Computational Class Field Theory • Computing Defining Equations • Cubic Number Fields • Ramification, Conductors, and Discriminants • Relative Class Groups, Units, and Regulators • Inverting Prime Ideals • Algorithms for *p*-adic fields

1999/APP. 576 PP., 7 ILLUS./HARDCOVER/$59.95
ISBN 0-387-98727-4
GRADUATE TEXTS IN MATHEMATICS, VOL. 193

JOEL L. SCHIFF, The University of Auckland, New Zealand

THE LAPLACE TRANSFORM

Theory and Applications

Contents: Basic Principles • Applications and Properties • Complex Variable Theory • Complex Inversion Formula • Partial Differential Equations

1999/APP. 256 PP., 65 ILLUS./HARDCOVER/$39.95
ISBN 0-387-98698-7
UNDERGRADUATE TEXTS IN MATHEMATICS

Springer
www.springer-ny.com

The Yellow Sale is now on!

If you have not yet received your copy, please call 1-800-SPRINGER and ask for catalog YS99, or visit the Yellow Sale website at www.springer-ny.com/yellowsale.
Sale ends 12/31/99.

JAMES M. HENLE, Smith College, Northhampton, MA and **TOM TYMOCZKO** (deceased)

SWEET REASON

A Field Guide to Modern Logic

Sweet Reason is a unique introductory text for courses on modern logic. Not only does this book cover all the basic rudiments of formal and informal logic, it also focuses students on the real world, where the discipline of logic adds substance and meaning to all kinds of human discourse. Everything from puzzles, paradoxes, and mathematical proofs to campaign debate excerpts, government regulations, and cartoons are used to show how logic is put to work by philosophers, mathematicians, advertisers, computer scientists, politicians, and others.

1999/APP. 672 PP., 25 ILLUS./SOFTCOVER
$39.95 (TENT.)/ISBN 0-387-98930-7
TEXTBOOKS IN MATHEMATICAL SCIENCES

GEORGE BACHMAN, Polytechnic University, Brooklyn, NY, **EDWARD BECKENSTEIN,** St. John's University, Staten Island, NY, and **LAWRENCE NARICI,** St. John's University, Jamaica, NY

FOURIER AND WAVELET ANALYSIS

Contents: Metric and Normed Spaces • Analysis • Bases • Fourier Series • The Fourier Transform • The Discrete and Fast Fourier Transforms • Wavelets

1999/APP. 512 PP., 31 ILLUS./HARDCOVER/$59.95
ISBN 0-387-98899-8
UNIVERSITEXT

Order Today!

Call: 1-800-SPRINGER or Fax: (201)-348-4505
Write: Springer-Verlag New York, Inc., Dept. S1511, PO Box 2485, Secaucus, NJ 07096-2485 · **Visit:** Your local technical bookstore
E-mail: orders@springer-ny.com · **Instructors:** Call or write for information on textbook exam copies

YOUR 30-DAY RETURN PRIVILEGE IS ALWAYS GUARANTEED!

11/99 Promotion Number S1511

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, DC 20036